

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number
WO 02/38744 A2(51) International Patent Classification⁷: **C12N 9/00**

(21) International Application Number: PCT/US01/51034

(22) International Filing Date: 18 October 2001 (18.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/241,573	18 October 2000 (18.10.2000)	US
60/243,643	25 October 2000 (25.10.2000)	US
60/245,256	2 November 2000 (02.11.2000)	US
60/248,395	13 November 2000 (13.11.2000)	US
60/249,826	16 November 2000 (16.11.2000)	US
60/252,303	20 November 2000 (20.11.2000)	US
60/250,981	1 December 2000 (01.12.2000)	US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LEE, Ernestine, A.** [US/US]; 624 Kains Street, Albany, CA 94706 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LAL, Preeti, G** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **YAO, Monique, G** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **WALIA, Narinder, K.** [US/US]; 890 David Street, #205, San Leandro, CA 94577 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **LEE, Sally** [US/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **GRIF-FIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **KALLICK, Deborah, A.** [US/US]; 900 Olive Street, Menlo Park, CA 94025 (US). **GANDHI,**

Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **ELILLOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue # 5D, San Francisco, CA 94122 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **NGUYEN, Danniell, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **LO, Terence, P.** [CA/US]; 1451 Beach Park Boulevard, Apt. 115, Foster City, CA 94404 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **PROTEASES**

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

WO 02/38744 A2

PROTEASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half-life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the

active catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) *Methods Enzymol.* 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaeobacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, *supra*).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and

prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J. Neurosci. 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) Adv. Neurol. 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all

eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) *Cell* 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, *supra*). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) *Annu. Rev. Med.* 50:57-74). A murine proto-oncogene, *Unp*, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) *Oncogene* 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) *Differentiation* 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) *J. Pathol.* 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) *Curr. Opin. Chem. Biol.* 3:584-591).

25 Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site

located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) *Methods Enzymol.* 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) *Arthritis Rheum.* 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) *J. Neurosci. Res.* 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, *supra*). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) *J. Neurotrauma* 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) *J. Neurol. Sci.* 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating

signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, supra; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, supra). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, supra; Thompson, C.B. (1995) Science 267:1456-1462).

20 Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the polypolyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2–3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum.

- Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) Crit. Rev. Oncol. 4:95-114).

Metalloproteases

- Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

- Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

- A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neurotensin (Serizawa, A. et al. (1995) J. Biol. Chem 270:2092-2098). Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) Neuropeptides 33:339-349).

- The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, L.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2} -cysteine interaction, or "cysteine switch," exposing

the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in Drosophila neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, supra). TACE has also been

identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) J. Biol. Chem. 272:556-562). To date eleven members are recognized by the Human Genome Organization (HUGO; <http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved>). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., *supra*) and/or procollagen processing (Colige, A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2374).

Protease inhibitors

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) Agents Actions Suppl. 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al. (1995) Biol. Biochem. Hoppe Seyler 376:71-80). Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin, that binds the proenzyme, proacrosin, and thereby aides in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) J. Biol. Chem. 269:10133-10140). The Kunitz family of serine protease inhibitors are characterized by one or more "Kunitz domains" containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter- α -trypsin inhibitor, and bikunin. (Marlor, C.W. et al. (1997) J. Biol. Chem. 272:12202-12208.) Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin. Aprotinin has clinical utility in reduction of perioperative blood loss.

The discovery of new proteases, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7," "PRTS-8," "PRTS-9," "PRTS-10," "PRTS-11," "PRTS-12," "PRTS-13," "PRTS-14," and "PRTS-15." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-15.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-15. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:16-30.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino

acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group
5 consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino
10 acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a
15 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an
20 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at
25 least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a
30 sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a
35 polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The

method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

35

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing

the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

5 "PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other
10 compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or
15 many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions,
20 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding
25 PRTS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example,
30 negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid
5 sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity
10 of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments
15 thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly
20 used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies
25 which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX
30 (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a
35 ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g.,

resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) *J. Biotechnol.* 74:5-13.)

5 The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) *Proc. Natl Acad. Sci. USA* 96:3606-3610).

 The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed
10 nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

 The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as
15 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
20 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

 The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"
25 refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

 "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,
30 3'-TCA-5'.

 A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be
35 employed as hybridization probes. The probes may be stored in freeze-dried form and may be

associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated
 5 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and
 10 assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded
 15 as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
20	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
25	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
30	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
35	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of
 40 the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative
5 polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or
10 absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be
15 assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a
20 fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected
25 from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:16-30 comprises a region of unique polynucleotide sequence that
30 specifically identifies SEQ ID NO:16-30, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:16-30 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:16-30 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:16-30 and the region of SEQ ID NO:16-30 to which the fragment corresponds are routinely
35 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-15 is encoded by a fragment of SEQ ID NO:16-30. A fragment of SEQ ID NO:1-15 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-15. For example, a fragment of SEQ ID NO:1-15 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-15. The precise length of a fragment of SEQ ID NO:1-15 and the region of SEQ ID NO:1-15 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>.

The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

5 *Matrix: BLOSUM62*
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 10 *Expect: 10*
 Word Size: 11
 Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,
 15 over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to
 25 the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

30 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default

residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 5 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

10 *Expect: 10*

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for 15 example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely 25 resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized 30 after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be 35 varied among experiments to achieve the desired stringency, and therefore hybridization specificity.

Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a

mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

5 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

10 The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

15 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

20 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

25 "Post-translational modification" of an PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

30 "Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target

DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization

technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

5 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a
10 recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is
15 expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

20 "Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear
25 sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a
30 cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular
35 structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding

molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

5 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

10 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

15 A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

25 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention

30

into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a

single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:3 is 50% identical to *Xenopus* ADAM 13 metalloprotease (GenBank ID g1916617) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.1\text{e-}208$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a neutral zinc metalloprotease active site domain and a disintegrin domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of these motifs is confirmed by BLIMPS, MOTIFS, and PROFILESCAN analyses, providing further corroborative evidence that SEQ ID NO:3 is a protease of the ADAM family. In an alternate example, SEQ ID NO:4 is 44% identical to human zinc metalloprotease ADAMTS7 (GenBank ID g5923788) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.2\text{e-}143$, which indicates the probability of obtaining the observed

polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a Reprolysin (M12B) family zinc metalloprotease site and a Thrombospondin type 1 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses

5 provide further corroborative evidence that SEQ ID NO:4 is a metalloprotease (note that the "Thrombospondin type 1 domains" are found at the carboxy-terminal end, and are characteristic of the ADAMTS metalloprotease protein family). In an alternate example, SEQ ID NO:5 is 62% identical to mouse distal intestinal serine protease (GenBank ID g5921501) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is

10 5.3e-99, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a trypsin family serine protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of this motif is confirmed by BLIMPS, MOTIFS, and PROFILESCAN analyses. BLIMPS analysis also reveals the

15 presence of kringle and type I fibronectin domains. Together, these data provide further corroborative evidence that SEQ ID NO:5 is a trypsin family serine protease. In an alternate example, SEQ ID NO:8 is 45% identical to human membrane-type serine protease 1 (GenBank ID g6002714) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.1e-69, which indicates the probability of obtaining the observed polypeptide

20 sequence alignment by chance. SEQ ID NO:8 also contains a trypsin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:8 is a serine protease. In an alternate example, SEQ ID NO:11 is 49% identical to mouse ADAM 4 protein precursor (GenBank ID g965014) as determined by the Basic Local Alignment Search Tool

25 (BLAST). (See Table 2.) The BLAST probability score is 4.1e-117, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also contains a reprolysin family propeptide domain and a disintegrin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of

30 conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is an ADAM family metalloprotease. In an alternate example, SEQ ID NO:12 is 42% identical to bovine enteropeptidase (GenBank ID g416132) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.2e-47, which indicates the probability of obtaining the observed

35 polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a trypsin domain as

determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is a trypsin family serine protease. In an alternate example, SEQ ID NO:13 is 52% identical from residues 110 to 482 to *Saccharomyces cerevisiae* Map1p methionine aminopeptidase (GenBank ID g662342) as determined by the Basic Local Alignment Search Tool (BLAST), with a probability score of 1.6e-99. (See Table 2.) SEQ ID NO:13 also contains a metallopeptidase family M24 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:13 is a methionine aminopeptidase. In an alternate example, SEQ ID NO:15 is 36% identical to *Xenopus* epidermis-specific serine protease (GenBank ID g6009515) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 7.7e-52, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a trypsin family protease active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of this motif is confirmed by BLIMPS, MOTIFS, and PROFILESCAN analyses. BLIMPS analysis also reveals that SEQ ID NO:15 contains a kringle domain, providing further corroborative evidence that SEQ ID NO:15 is a protease of the trypsin family. SEQ ID NO:2-3, SEQ ID NO:6-7, SEQ ID NO:9-10 and SEQ ID NO:14 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-15 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:16-30 or that distinguish between SEQ ID NO:16-30 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5')

and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7635792H1 is the identification number of an Incyte cDNA sequence, and SINTDIE01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 55147856J1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g876900) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3,...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX_gAAAAA_gBBBBB_1_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/ r examples f programs
--------	---

5	GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
	GBI	Hand-edited analysis of genomic sequences.
	FL	Stitched or stretched genomic sequences (see Example V).
	INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:16-30, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting

of SEQ ID NO:16-30. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding PRTS. A splice variant may have portions which have
5 significant sequence identity to the polynucleotide sequence encoding PRTS, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS over its entire
10 length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding PRTS. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

15 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These
20 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected
25 conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide
30 sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the
35 synthetic sequence may be inserted into any of the many available expression vectors and cell

systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID
5 NO:16-30 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of
10 the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is
15 automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences
20 are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,
25 such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising
30 a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region
35 of unknown sequence before performing PCR. Other methods which may be used to retrieve

unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed
5 using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been
10 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze
15 the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire
20 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS,
25 or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but
30 not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or
5 improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of
10 DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby
15 maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)
20 Alternatively, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the
25 amino acid sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)
30 The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in
35 a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and

inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where
 5 sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various
 10 origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control
 15 elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences
 20 encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or
 25 animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and
 30 Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al.

(1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)
The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc.

Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
5 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression
10 of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance
15 to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine
20 phosphoribosyltransferase genes, for use in *tk* and *ap^r* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,
25 Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate
30 luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the
35 sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing

sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

5 In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

10 Immunological methods for detecting and measuring the expression of PRTS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a
15 competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

20 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector
25 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for
30 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence
35 and/or the vector used. As will be understood by those of skill in the art, expression vectors

containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

5 In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the
10 compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PRTS or the
15 compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound.

20 Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds
25 that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a
30 compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

30 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with reproductive, normal and tumorous gastrointestinal, urogenital, bone tumor, breast, brain, testis, and adrenal tumor tissues, as well as with adherent mononuclear cells. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative,

developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

- 5 Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis,
- 10 gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic
- 15 obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha₁-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-
- 20 occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular
- 25 replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus,
- 30 carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune
- 35 polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis,

5 myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and

10 trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall

15 bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-

20 Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema,

25 allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis

30 dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions,

35 telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug

reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases,
 epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic
 palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et
 plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's
 5 corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal
 nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic
 hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy,
 ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,
 Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic
 10 lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis
 pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and
 viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial
 thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases
 including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal
 15 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,
 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental
 retardation and other developmental disorders of the central nervous system including Down
 syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve
 disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral
 20 nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and
 toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety,
 and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia,
 diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia,
 Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial
 25 frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease,
 ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous
 cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation
 syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic
 pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a
 30 disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the
 prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the
 male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative
 thereof may be administered to a subject to treat or prevent a disorder associated with decreased
 35 expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

- 5 In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

- In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such
10 disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

- 15 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

- In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate
20 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

- 25 An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and
30 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

- For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to
35 increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral

gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to
5 PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

10 Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and
15 Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda,
20 S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

25 Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated.
30 For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its
5 specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an
10 association constant, K_a , which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a
15 particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the
20 antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For
25 example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

30 In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well known in the art, and antisense oligonucleotides or larger fragments

can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing

these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining
5 retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-
10 7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to
15 the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby
20 incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with
25 respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.
30 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this
35 patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22.

For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of
5 herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based
10 on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PRTS into the
15 alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy
20 application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the
25 art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of
30 polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze
5 endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for
10 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques
15 for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines,
20 cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs
25 and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a
30 compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective
35 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of

polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treatment of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved

using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and
5 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of
10 Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal,
15 enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides
20 and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active
25 ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of
30 the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell
35 culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs,

monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS
5 or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the
10 therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the
15 patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the
20 subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and
25 methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

30 In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body
35 fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and

may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values.

10 Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

15

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

20

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:16-30 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

25

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

30

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, 5 emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic 10 obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, α_1 -antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno- 15 occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular 20 replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, 25 carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune 30 polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, 35 myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis,

pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and

5 trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall

10 bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-

15 Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema,

20 allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis

25 dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions,

30 telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's

35 corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal

nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a

suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are

substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to
5 amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-
10 throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence
15 chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C.
20 et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the
25 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor
30 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her
35 pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to
5 generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by
10 hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

15 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of
20 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test
25 compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes
30 are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of
35 Environmental Health Sciences, released February 29, 2000, available at

<http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a

variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either

coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a
5 chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop
10 genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic
15 map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as
20 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely
25 localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

30 In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below including U.S. Ser. No. 60/241,573, U.S. Ser. No. 60/243,643, U.S. Ser. No. 60/245,256, U.S. Ser. No. 60/248,395, U.S. Ser. No. 60/249,826, U.S. Ser. No. 60/252,303, U.S. Ser. No. 60/250,981, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was
5 isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the
10 recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column
15 chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte
20 Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo
25 excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1
30 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically

using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

- 5 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cyclers or the PTC-200 thermal cyclers (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as
- 10 the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading
- 15 frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

- The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and
- 20 programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae,
 - 25 Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to
 - 30 produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length
 - 35 polynucleotide sequences were translated to derive the corresponding full length polypeptide

sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov
5 model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS.PRO software, (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also
10 calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate
15 references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide
20 and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:16-30. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteases were initially identified by running the Genscan gene identification
25 program against public genomic sequence databases (e.g., gbpr and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a
30 FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were
35 then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the

Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST

analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:16-30 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:16-30 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is

much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

5 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA
10 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

 In like manner, full length polynucleotide sequences are verified using the above procedure or
15 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

 Hybridization probes derived from SEQ ID NO:16-30 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a
25 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon
30 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

35 X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

5 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia
10 Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water,
15 and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic
20 apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in
25 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered
30 with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X
35 SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is
5 focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.
10 Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,
15 although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples
20 from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital
25 (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping
30 emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

10 XII. Expression of PRTS

Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman

Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII, and XIX where applicable.

5 XIII. Functional Assays

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which
 10 contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the
 15 recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events
 20 include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of
 25 fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human
 30 immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

35 XIV. Production of PRTS Specific Antibodies

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. Chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) Anal. Biochem. 247:305-309).

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved; and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is

introduced on an inducible vector so that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) FEBS Lett. 447:53-57).

XVIII. Identification of PRTS Substrates

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

To screen for in vivo PRTS substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT 10-3 Phage display vector, Novagen, Madison WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PRTS Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific

embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6926819	1	6926819CD1	16	6926819CB1
7473526	2	7473526CD1	17	7473526CB1
7478443	3	7478443CD1	18	7478443CB1
3533147	4	3533147CD1	19	3533147CB1
7483438	5	7483438CD1	20	7483438CB1
7246467	6	7246467CD1	21	7246467CB1
7997881	7	7997881CD1	22	7997881CB1
7484378	8	7484378CD1	23	7484378CB1
7473143	9	7473143CD1	24	7473143CB1
4382838	10	4382838CD1	25	4382838CB1
6717888	11	6717888CD1	26	6717888CB1
7472044	12	7472044CD1	27	7472044CB1
7477384	13	7477384CD1	28	7477384CB1
7077175	14	7077175CD1	29	7077175CB1
7480124	15	7480124CD1	30	7480124CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO: or PROTEOME ID NO:	Probability Score	Annotation
1	6926819	g190418	2e-143	[Homo sapiens] preprocathepsin L precursor (Joseph, L.J. et al. (1988) J. Clin. Invest. 81 (5), 1621-1629)
2	7473526	g4481747	1.6e-102	[Rattus norvegicus] calpain Ri88 (Shearer, T.R. et al. (2000) Methods Mol Biol. 144:277-85)
3	7478443	g13157560	0.0	[3' incom][Homo sapiens] dJ964F7.1 (novel disintegrin and reprolysin metalloproteinase family protein)
3	7478443	g1916617	2.1e-208	[Xenopus laevis] ADAM 13 (Alfandari, D. et al. (1997) Dev. Biol. 182 (2), 314-330)
4	3533147	g5923788	2.2e-143	[Homo sapiens] zinc metalloprotease ADAMTS7 (Hurskainen, T.L. et al. (1999) J. Biol. Chem. 274 (36), 25555-25563)
5	7483438	g5921501	5.3e-99	[Mus musculus] distal intestinal serine protease (Shaw-Smith, C.J. et al. (2000) Biochim. Biophys. Acta 1490 (1-2), 131-136)
6	7246467	g9971757	1.3e-182	[Homo sapiens] (AF229438) ubiquitin-specific processing protease (Kim, J. et al. (2000) Genome Res. 10 (8), 1138-1147)
6	7246467	g13603869	1e-179	[fl][Homo sapiens] ubiquitin specific protease 26 (Wang, P.J. et al, (2001) Nat. Genet. 27 (4), 422-426)
7	7997881	g2739431	3.9e-94	[Mus musculus] hematopoietic-specific IL-2 deubiquitinating enzyme (Zhu, Y. et al. (1997) J. Biol. Chem. 272 (1), 51-57)
8	7484378	g6002714	6.1e-69	[Homo sapiens] membrane-type serine protease 1 (Takeuchi, T. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96 (20), 11054-11061)
9	7473143	g10185056	4e-23	[Gallus gallus] (AJ012462) colloid protein (tollid-related metalloprotease) (Liaubet, L. et al. (2000) Mech. Dev. 96 (1), 101-105)
9	7473143	g14794726	1e-131	[fl][Homo sapiens] CUB and sushi multiple domains 1 protein (Sun, P.C. et al. (2001) Genomics. 75 (1-3), 17-25)
10	4382838	g4929103	3.7e-19	[Hydra vulgaris] metalloproteinase 2 (Yan, L. et al. (2000) Development 127 (1), 129-141)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO: or PROTEOME ID NO:	Probability Score	Annotation
11	6717888	g965014	4.1e-117	[Mus musculus] ADAM 4 protein precursor (Wolfsberg, T.G. et al. (1995) Dev. Biol. 169 (1), 378-383)
11	6717888	g6110345	3.00e-78	[fl][Homo sapiens] metalloproteinase-disintegrin beta (Cerretti, D.P. et al. (1999) Biochem. Biophys. Res. Commun. 263 (3), 810-815)
12	7472044	g416132	2.20e-47	[Bos taurus] enteropeptidase (LaVallie, E.R. et al. (1993) J. Biol. Chem. 268 (31), 23311-23317)
13	7477384	g662342	1.60e-99	[Saccharomyces cerevisiae] Map1p: methionine aminopeptidase
14	7077175	g9757702	1.40e-54	[Xenopus laevis] homolog of human MT-SP1 (Yamada, K. et al. (2000) Gene 252 (1-2), 209-216)
15	7480124	g6009515	7.7e-52	[Xenopus laevis] epidermis specific serine protease (Yamada K et al. (1999) Dev Biol. 214(2):318-30)
15	7480124	g13516326	4e-52	[fl][Homo sapiens] marapsin

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6926819CD1	334	S161 T116 T156 T211	N222	Papain family cysteine protease Peptidase_C1: T114-T333 Eukaryotic cysteine protease active sites thiol protease_cys.prf: E113-E164 thiol protease_his.prf: S253-G308 EUKARYOTIC THIOL PROTEASES CYSTEI DM00081 P07711 19-332: L19-V334 DM00081 P25975 20-333: D22-V334 DM00081 P06797 19-332: D22-V334 DM00081 P15242 20-332: T20-V334 PROTEASE PRECURSOR SIGNAL CYSTEINE THIOL ZYMOGEN CATHEPSIN GLYCOPROTEIN PD000158: Y190-P332, T116-S219 Eukaryotic thiol protease active site BL00139: Q133-F142, N176-M184, D276-S285, Y296-Y312 PAPAIN CYSTEINE PROTEASE PR00705: Q133-L148, H277-E287, Y296-S302 Eukaryotic cysteine proteases active sites Thiol_Protease_Asn Y296-M315 Thiol_Protease_Cys Q133-A14 Thiol_Protease_His L275-S285 signal_peptide: M1-T20 signal_cleavage: M1-A17	HMME PFAM PROFILES SCAN BLAST_DOMO BLIMPS_BLOCKS BLIMPS_PRINTS MOTIFS HMME SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	7473526CD1	511	S126 S18 S188 S22 S293 S294 S300 S393 S449 S508 T128 T168 T265 T319 T351 T362	N124 N231	Calpain family cysteine protease Peptidase_C2: L43-T336 Calpain large subunit, domain III Calpain_III: K347-S490, Eukaryotic cysteine protease active sites BL00139: Q95-L104, L273-W289 CALPAIN CYSTEINE PROTEASE PR00704: L162-L185, G187-L214, N312-C333, A363-F380, Q28-A51, W71-L93, Q95-A111, Y131-V156 CALPAIN CATALYTIC DOMAIN DM01305 P17655 1-505: Q28-R482 DM01305 P20807 19-581: Q27-G242, R232-P480 DM01305 S57196 12-574: Y17-I235, G249-P480 DM01305 A48764 1-507: Q28-P480 PROTEASE CALPAIN HYDROLASE SUBUNIT NEUTRAL THIOL LARGE CALCIUMACTIVATED PROTEINASE PD001545: L43-C237, A176-T336 PD001874: K347-P480 Eukaryotic cysteine proteases active site Thiol_Protease_Cys: Q95-A106	HMMER_PFAM HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_DOMO BLAST_PRODUM MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7478443CD1	812	S162 S389 S450 S547 S55 S61 S639 S787 T174 T208 T258 T264 T302 T605 Y243	N109 N145 N231 N276 N448	<p>Reprolysin family propeptide Pep_M12B_propep:\nE80-Q198</p> <p>Reprolysin (M12B) family zinc metalloprotease Reprolysin: K210-P409</p> <p>Neutral Zn metalloprotease, Zn-binding region zinc_protease.prf: E323-A375</p> <p>Neutral Zn metalloprotease, Zn-binding region BL00142: T342-G352</p> <p>Neutral Zn metalloprotease, Zn-binding region Zinc_Protease T342-L351</p> <p>do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368 S60257 204-414: R202-D410 DM00368 Q05910 189-395: R206-D410 DM00591 S60257 492-628: F487-G608 DM00368 P28891 1-202: E204-P409</p> <p>METALLOPROTEASE PRECURSOR HYDROLASE SIGNAL ZINC VENOM CELL TRANSMEMBRANE ADHESIO PD000791: R209-P409</p> <p>CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGN PD000664: E426-Y500</p> <p>PRECURSOR METALLOPROTEASE SIGNAL CELL ZINC HYDROLASE TRANSMEMBRANE ADHESION PROTEIN PD000935: L70-M169</p>	<p>HMMER_PFAM</p> <p>HMMER_PFAM</p> <p>PROFILES SCAN</p> <p>BLIMPS_BLOCKS</p> <p>MOTIFS</p> <p>BLAST_DOMO</p> <p>BLAST_PROD OM</p> <p>BLAST_PROD OM</p>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					TRANSMEMBRANE METALLOPROTEASE SIGNAL PRECURSOR GLYCOPROTEIN CELL FERTILIN BETA ADHESION PD001269: D503-L572 Disintegrin: E426-L501 Disintegrins proteins BL00427: C443-P497 DISINTEGRIN SIGNATURE PR00289: C457-R476, E486-D498 Disintegrins signature disintegrins.prf: G352-D498 signal_peptide: M1-G27 signal_cleavage: M1-G27	BLAST_PRODOM HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS PROFILES SCAN HMMER SPSCAN
4	3533147CD1	1236	S1031 S1096 S1104 S1108 S1219 S164 S201 S297 S30 S406 S55 S61 S620 S641 S650 S670 S708 S786 S800 S834 T10 T1083 T143 T222 T235 T391 T400 T41 T431 T644 T684 T827 T875 T904 T968 T995 Y191 Y516	N323 N44 N754 N793 N848 N918 N948	Reprolysin (M12B) family zinc metalloprotease V305-P508 Thrombospondin type 1 domain: S603-C653, G1140-C1192, W1067-P1113	HMMER_PFAM HMMER_PFAM

Table 3 (cont.)

[illegible]

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8					Serine proteases, trypsin family, active sites trypsin_his.prf: L590-Q963 Serine proteases, trypsin family, active sites trypsin_ser.prf: L435-R1105	PROFILES SCAN
9	7473143CD1	462	S194 S226 S308 S392 S438 S50 S53 S56 T170 T318 T372 T448	N138 N147 N164	CLR/C1S REPEAT DM00162 A55362 33-151:C315-Y416 DM00162 P98069 303-417:A314-Q417 DM00162 JH0403 32-150:C315-Y416 DM00162 A57190 826-947:C315-Y416, C139-Y244 GLYCOPROTEIN DOMAIN EGF-LIKE PROTEIN PRECURSOR SIGNAL RECEPTOR INTRINSIC FACTOR B12 REPEAT PD000165:C315-V418, N147-V246 Alpha-lytic endopeptidase serine protease (S2A) signature PR00861:K78-S92 CUB domain CUB: C139-Y244, C315-Y416	BLAST_DOMO
10	4382838CD1	659	S123 S181 S263 S4 S616 S656 S86 T103 T171 T202 T27 T330 T432 T462 T506 T564 T92 T94 Y118	N137 N146 N207 N313 N406	MAM domain: F453-K624 Immunoglobulin domain: G55-A122, G161-T220, D257-I316, C459-S541 MAM domain proteins BL00740: C459-W471, E607-L627 MAM domain signature PR00020A: N457-S475, M502-L518, Y538-Q549, V586-G600, G605-A618 PRECURSOR GLYCOPROTEIN SIGNAL TRANSMEMBRANE HYDROLASE PROTEIN REPEAT RECEPTOR PHOSPHATASE NEUROFILIN PD001482:F453-K624	BLIMPS_PRINTS BLIMPS_PRINTS BLAST_PRODOME

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10					MAM DM01344 P28824 595-796:T462-D613 DM01344 P98072 352-509:F453-D614 DM01344 A55620 618-796:D465-G608	BLAST_DOMO
11	6717888CD1	626	S152 S184 S279 S299 S323 S329 S407 T127 T175 T192 T305 T38 T597 T91	N224 N405 N529	Signal cleavage: M1-C28 Signal peptide: M1-C28 Transmembrane domain: N596-V613 Reprolysin family propeptide domain: H75-E191 Disintegrin domain: N314-I389 Disintegrins signature: G322-D386 DISINTEGRIN SIGNATURE PR00289: C345-R364, E374-D386 PRECURSOR METALLOPROTEASE SIGNAL CELL HYDROLASE TRANSMEMBRANE PROTEASE ADHESION PROTEIN ZINC PD000935:F37-V158 CELL ADHESION PLATELET BLOOD COAGULATION VENOM DISINTEGRIN METALLOPROTEASE PRECURSOR SIGNAL PD000664:E317-Y388 TRANSMEMBRANE PRECURSOR SIGNAL FERTILIN GLYCOPROTEIN PROTEIN BETA METALLOPROTEASE CELL INTEGRIN PD001734:R528-R599 do ZINC; NEUTRAL; METALLOPEPTIDASE; HEMORRHAGIC; DM00533 S59854 14-197:L19-R194 DM00533 I48101 14-195:L19-E164	SPSCAN HMMER HMMER HMMER_PFAM HMMER_PFAM PROFILES SCAN BLIMPS_PRINTS BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					Go ZINC; REGULATED; EPIDIDYMAL; NEUTRAL; DM00591 I48101 475-621:Y375-L526 DM00591 S55059 511-662:Y375-C530	BLAST_DOMO
12	7472044CD1	557	S120 S138 S180 S282 S315 S320 S62 T133 T229 T265	N65	Signal cleavage: M1-G45 Signal peptide: M1-G19 Trypsin domain: I73-L306 Serine proteases, trypsin family, active sites BL00134: P293-L306, C98-C114, D254-L277 Type I fibronectin domain BL01253: C98-A111, T173-E209, V253-C266, E275-Q309 Kring domain proteol BL00021: C98-F115, V184-G205, T265-L306 Serine proteases, trypsin family, active sites: L90-E145, S241-E289 Chymotrypsin serine proteases signature PR00722: G99-C114, T161-V175, V253-T265 PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: E136-L306, I73-R231 TRYPSIN DM00018 P98072 800-1033:R72-Q307 DM00018 P20918 576-808:G71-M310 DM00018 P26262 391-624:I73-Q309 DM00018 P05981 163-403:I73-L306	SPSCAN HMMER HMMER_PFAM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS PROFILESAN BLIMPS_PRINTS BLAST_PRODOM BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12					Serine proteases, trypsin family, histidine active site: 109-C114 Serine proteases, trypsin family, serine active site: 254-T265	MOTIFS
13	7477384CD1	494	S224 S240 S304 S485 S92 T155 T192 T270 T359 T447 T474 T480	N341	Metalloprotease family M24 domain: 237-E477 Methionine aminopeptidase signature: 379-M437 Methionine aminopeptidase signature PR00599: 301-P314, D323-D339, F393-H405, 424-P436 AMINOPEPTIDASE METHIONINE PRECURSOR METAP PEPTIDASE M MAP HYDROLASE COBALT PUTATIVE D035886:C117-Q228 AMINOPEPTIDASE HYDROLASE METHIONINE PEPTIDASE PROTEIN COBALT M DIPEPTIDASE XPRO MAP D000555:I236-F393 METHIONINE AMINOPEPTIDASE M01530 Q01662 123-375:S234-R482 M01530 P53579 1-252:I236-T480 M01530 P07906 1-252:I236-D484 M01530 P44421 1-253:I236-R482	HMMER_PFAM PROFILES SCAN BLIMPS_PRINTS BLAST_PROD OM BLAST_PROD OM
14	7077175CD1	593	S39 S66 S118 S150 S273 S418 S508 T65 T120 T188 T212 T302 T393 T420 T500 T548 T570	N35 N300 N391 N416 N539	Trypsin: 94-I184, V257-I484	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14					Serine protease, trypsin family active site trypsin_his.prf: 274-S322 trypsin_ser.prf: L119-E167, L419-Q467	PROFILESCAN
					Serine protease, trypsin family active site B00134: C282-C298	BLIMPS_BLOCKS
					Serine protease, trypsin family histidine active site: L293-C298 serine active site: D134-V145, D434-A445	MOTIFS
					Chymotrypsin serine protease family (S1) PR00722C: V133-V145	BLIMPS_PRINTS
					TRYPSIN DM00018 P06872 24-242: V257-I484, P95- I184 DM00018 P00762 24-242: V257-I484, V92- I184 DM00018 P07146 24-242: V257-I484, V92- I184 DM00018 S13813 24-242: V257-E486, P95- E186	BLAST_DOMO
					Kringles domain proteins BL00021: C282-F299	BLIMPS_BLOCKS
					Type I fibronectin domain BL01253: V133-C146	BLIMPS_BLOCKS
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE PD0000046: Y337-I484, P95-I184, V257- T393	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7480124CD1	319	T190 T232	N118 N170 N247	<p>Trypsin: I53-I281</p> <p>Serine protease, trypsin family active site</p> <p>trypsin_ser.prf: I217-N264</p> <p>trypsin_his.prf: V70-H119</p> <p>Serine proteases, trypsin family serine active site: T232-V243</p> <p>histidine active site: L89-C94</p> <p>Serine protease, trypsin family active sit</p> <p>BL00134: C78-C94, T232-I255, P268-I281</p> <p>Chymotrypsin serine protease family (S1) PR00722: G79-C94, P139-I153, K231-V243</p> <p>TRYPSIN</p> <p>DM00018 P03951 389-621: L54-D283</p> <p>DM00018 P14272 391-624: I53-I</p> <p>DM00018 A57014 45-284: I53-I2</p> <p>DM00018 P26262 391-624: I53-D283</p> <p>Kringe domain proteolins</p> <p>BL00021: C78-F95, G240-I281</p> <p>PROTEASE SERINE PRECURSOR SIGNAL</p> <p>HYDROLASE ZMOGEN GLYCOPROTEIN FAMILY MULTIGENE</p> <p>PD000046: Q121-I281, I53-S191</p> <p>transmembrane_domain: P304-L320</p> <p>signal_cleavage: M1-A25</p> <p>signal peptide: M1-A25</p>	<p>HMMER_PFAM</p> <p>PROFILES SCAN</p> <p>MOTIFS</p> <p>BLIMPS_BLOCKS</p> <p>BLIMPS_PRINTS</p> <p>BLAST_DOMO</p> <p>BLIMPS_BLOCKS</p> <p>BLAST_PRODOM</p> <p>HMMER</p> <p>SPSCAN</p> <p>HMMER</p>

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16	6926819CB1	2406	1530-1554, 1- 579, 2358-2406	7635792H1 (SINTDIE01)	1192	1427
				6926819H1 (SINTMR01)	1	584
				55147856J1	289	1122
				g876900	640	1167
				GB1.g10635543_ 2.edit	1573	2406
17	7473526CB1	1967	668-1471, 1815- 1967, 34-628	72474147D1	1	800
				72473150D1	704	1360
				7600172R6 (ESOGTME01)	1233	1967
				6431466H1 (LUNGNON07)	933	1404
18	7478443CB1	3446	1-2009, 2548- 2726	6603789H1 (UTREDIT07)	1733	2275
				766351911 (UTRSTME01)	1245	1890
				7686903H1 (PROSTME06)	563	1125
				8008540H1 (NOSEDIC02)	1464	2014
				7174969F8 (BRSTTMC01)	637	1280
				58002846H1	2115	2926
				5600214H1 (UTRENON03)	1	584

Table 4

Polynucleotide SEQ ID NO:	Incye Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
18				2880287T6 (UTRSTUT05)	2852	3446
19	3533147CBI	4888	3526-4295, 1- 1893, 2225-3447	72335984V1	3933	4760
				55124332HI	1836	2554
				3533147T6 (KIDNNOT25)	4165	4805
				58002730I1	2744	3447
				72024457V1	3283	4091
				58002714J1	2220	3133
				55054505J1	880	1878
				2503829F6 (CONUTUT01)	4539	4888
				55054461J2	1164	1880
				55064725H1	605	1085
				GNN.g7710798_ 000007_004.edit	803	1352
				GNN.g9256175_ 000001_010.edit	1	806
20	7483438CBI	1074	1-561	g2114954	310	562
				999322R6 (KIDNTUT01)	627	1074
				g2054575	563	949
				60201955V1	563	746
				GBI.g2734091_0 00001.edit	1	1062

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20				2108692H1 (BRAITUT03)	1	258
				ENST000000055	144	547
				11		
21	7246467CB1	3573	2035-2345, 3547- 3573, 948-1045	55049833J1	2364	3214
				g1401740	368	969
				g1401739	1	643
				7409686H1	2973	3573
				(BRAIFEJ02)		
				GB1:g8980973_0	1818	2387
				00025_000020_0		
				00026.edit		
				GB1:g9739342_0	498	1625
				00015.edit		
				7246467F8	1302	1961
				(PROSTMY01)		
				4439185H1	313	473
				(SINTNOT22)		
				4325701H1	2258	2538
				(TLYMUNT01)		
22	7997881CB1	4659	1-111, 676-937, 3045-4659	72470070D1	696	1428
				72474695D1	1458	2168
				55136785H1	481	1231
				g1479172	83	570
				71390157V1	3966	4659

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
22	7997881CB1	4659	1-111, 676-937,	72470070D1	696	1428
			1-111, 676-937, 3045-4659	72474695D1	1458	2168
			1-111, 676-937, 3045-4659	55136785H1	481	1231
			1-111, 676-937, 3045-4659	g1479172	83	570
			1-111, 676-937, 3045-4659	71390157V1	3966	4659
			1-111, 676-937, 3045-4659	GBI:g10434351	88	3747
			1-111, 676-937, 3045-4659	70151956V1	83	546
			1-111, 676-937, 3045-4659	71194256V1	3890	4592
			1-111, 676-937, 3045-4659	g4372794	3424	3858
			1-111, 676-937, 3045-4659	72473269D1	1366	2047
			1-111, 676-937, 3045-4659	6753547H1 (SINTFER02)	3432	3952
			1-111, 676-937, 3045-4659	72473189D1	1742	2488
			1-111, 676-937, 3045-4659	7997881H1	1	625

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
23	7484378CB1	3711	324-1357, 1-141, 1533-1676, 1814- 1992, 2823- 3224, 2283-2671	7413053H1 (BONMTUE02)	3294	3711
				GNN.g6015230_ 000111_006	2260	3387
				GNN.g6015210_ 000057_004.edit	1	1678
				GNN.g6015230_ 000110_002	1223	2259
				55147453J1	1868	2429
24	7473143CB1	2017	1-1730, 1998- 2017	72342184D1	1273	2017
				GNN.g6778515_ 000015_002	485	1043
				6987935F8 (BRAIFER05)	1	597
				72341987D1	851	1664
25	4382838CB1	2646	2501-2646, 2106- 2161, 1-120, 811- 1102, 2278-2316	g764817	2162	2646
				3145451R7 (HNT2AZS07)	1651	2307
				72611602V1	1166	1790
				8463589U1	1908	2317
				72611354V1	678	1405

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
25				7114350R6 (BRAENOK01)	1	667
				72481694D1	521	1195
26	6717888CB1	2088	1-111, 676-937, 3045-4659	6717888F6 (CONDUTUT02)	1232	2026
				55072203H1	1	219
				5801608F8 (BONRFET03)	914	1428
				55047486J1	66	598
				1506340H1 (BRAITUT07)	1877	2088
				6247581F6 (TESTNOT17)	448	925
27	7472044CB1	1890	799-1363, 1-759, 1549-1890	2499087F6 (ADRETUT05)	1611	1890
				FL405947_00001	1	1872
28	7477384CB1	2984	1-410, 1990- 2021	71346663V1	2330	2984
				GNN_g9293863_4	602	1377
				GBL_g9293863.ed	1	557
				it		
				8325462U1	355	1064
				70684193V1	1724	2436
				8450123U1	830	1719
				71346028V1	1748	2465
				70683404V1	1156	1747

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
29	7077175CB1	2255	782-1241, 1-606, 1532-1986, 2209- 2255	7077175F8 (BRAUTDR04)	928	1099
				GNN_1311	198	1508
				GBI_edit_2	1	336
				55147453J1	1117	1678
				GBI_edit_1	1509	2255
30	7480124CB1	1250	1226-1250	g2051416	711	1250
				56009032J1	1	646
				g2057296	512	1093

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
16	6926819CB1	SINTDIE01
17	7473526CB1	ESOGTME01
18	7478443CB1	UTRSTME01
19	3533147CB1	CONUTUT01
20	7483438CB1	KIDNTUT01
21	7246467CB1	TESTNOT03
22	7997881CB1	BRSTNOT07
23	7484378CB1	BONMTUE02
24	7473143CB1	PONSAZT01
25	4382838CB1	BRAENOK01
26	6717888CB1	TESTNOT17
27	7472044CB1	ADRETUT05
28	7477384CB1	MPHGNOT03
30	7077175CB1	BONSTUT01

Table 6

Library	Vector	Library Description
ADRETUT05	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
BONNTUE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female during an exploratory laparotomy with soft tissue excision. Pathology indicated giant cell tumor of the sacrum. The patient presented with pelvic joint pain, constipation, urinary incontinence, and unspecified abdominal/pelvic symptoms. Patient history included a soft tissue malignant neoplasm. Patient medication included Darvocet. Family history included prostate cancer in the grandparent(s).
BONSTUT01	pINCY	Library was constructed using RNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female during an exploratory laparotomy with soft tissue excision. Pathology indicated giant cell tumor of the sacrum. Patient history included a soft tissue malignant neoplasm. Family history included prostate cancer.
BRAENOK01	PSPORT1	This amplified and normalized library was constructed using RNA isolated from inferior parietal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were multiple small microscopic areas of cavitation with surrounding gliosis scattered throughout the cerebral cortex. Patient history included dilated cardiomyopathy, congestive heart failure, and cardiomegaly. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec. 1.08 million independent clones from this amplified library were normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRSTNOT07	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.

Table 6 (cont.)

Library	Vector	Library Description
CONUTUT01	pINCY	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed müllerian tumor present in the sigmoid mesentery at two sites.
ESOGTME01	PSPORT1	This 5' biased random primed library was constructed using RNA isolated from esophageal tissue removed from a 53-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, and regional lymph node biopsy. Pathology indicated no significant abnormality in the non-neoplastic esophagus. Pathology for the matched tumor tissue indicated invasive grade 4 (of 4) adenocarcinoma, forming a sessile mass situated in the lower esophagus, 2 cm from the gastroesophageal junction and 7 cm from the proximal margin. The tumor invaded through the muscularis propria into the adventitial soft tissue. Metastatic carcinoma was identified in 2 of 5 paragastric lymph nodes with perinodal extension. The patient presented with dysphagia. Patient history included membranous nephritis, hyperlipidemia, benign hypertension, and anxiety state. Previous surgeries included an adenotonsillectomy, appendectomy, and inguinal hernia repair. The patient was not taking any medications. Family history included atherosclerotic coronary artery disease, alcoholic cirrhosis, alcohol abuse, and an abdominal aortic aneurysm rupture in the father; breast cancer in the mother; a myocardial infarction and atherosclerotic coronary artery disease in the sibling(s); and myocardial infarction and atherosclerotic coronary artery disease in the grandparent(s).
KIDNTUT01	PSPORT1	Library was constructed using RNA isolated from the kidney tumor tissue removed from an 8-month-old female during nephroureterectomy. Pathology indicated Wilms' tumor (nephroblastoma), which involved 90 percent of the renal parenchyma. Prior to surgery, the patient was receiving heparin anticoagulant therapy.
MPHGNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.
PONSAZT01	pINCY	Library was constructed using RNA isolated from diseased tons tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

Table 6 (cont.)

Library	Vector	Library Description
SINTDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from small intestine tissue removed from a 49-year-old Caucasian female during gastroenterostomy, exploratory laparotomy, and vagotomy. The patient presented with acute stomach ulcer with obstruction, nausea and vomiting, and abnormal weight loss. Patient history included backache, acute stomach ulcer with perforation, and normal delivery. Previous surgeries included adenotonsillectomy and total abdominal hysterectomy. Patient medications included Premarin. Family history included benign hypertension, type II diabetes and congestive heart failure in the father.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
TESTNOT17	PINCY	Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
UTRSTWE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from uterus tissue removed from a 49-year-old Caucasian female during vaginal hysterectomy and bilateral salpingo-oophorectomy. Pathology for the matched tumor tissue indicated multiple (6) intramural leiomyomata. The patient presented with excessive menstruation, deficiency anemia, and dysmenorrhea. Patient history included abdominal pregnancy, headache, and chronic obstructive asthma. Previous surgeries included hemorrhoidectomy, knee ligament repair, and intranasal lesion destruction. Patient medications included Azmacort, Proventil, Trazadone, Zostrix HP, iron, Premarin, and vitamin C. Family history included alcohol abuse, atherosclerotic coronary artery disease, upper lobe lung cancer, and carotid endarterectomy in the father; breast fibroadenosis in the sibling(s); and acute myocardial infarction, liver cancer, acute leukemia, and breast cancer (central) in the grandparent(s).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved-Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	

Table 7

Program	Description	Reference	Parameter Threshold
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

5

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- 10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,
b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,
c) a polynucleotide complementary to a polynucleotide of a),
15 d) a polynucleotide complementary to a polynucleotide of b), and
e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

20

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

30

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

5

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence
10 selected from the group consisting of SEQ ID NO:1-15.

19. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 17.

15

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of
25 functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

30

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

35

25. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
 - b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

10

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

25

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

30

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions

35

whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

5 d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

10 30. A diagnostic test for a condition or disease associated with the expression of PRTS in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

15 b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

20 b) a single chain antibody,

c) a Fab fragment,

d) a F(ab')₂ fragment, or

e) a humanized antibody.

25 32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

30

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of PRTS in a subject, comprising administering to said subject an effective amount of the composition of claim

35 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
5 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

10 52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

15 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical
20 location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

25 56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

30 59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

35 61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
- 5 64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
- 10 67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
- 15 69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20 NO:16.
72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:17.
- 25 73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:18.
74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:19.
- 30 75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:20.
76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
35 NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
5 NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:24.

10 80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:26.
15

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20 NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:29.

25 85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:30.

<110> INCYTE GENOMICS, INC.
 LEE, Ernestine A.
 HAFALIA, April
 YUE, Henry
 LAL, Preeti G.
 YAO, Monique G.
 LU, Yan
 WALIA, Narinder K.
 WARREN, Bridget A.
 LU, Dyung Aina M.
 BAUGHN, Mariah R.
 DELEGEANE, Angelo M.
 BURFORD, Neil
 BOROWSKY, Mark L.
 LEE, Sally
 XU, Yuming
 GRIFFIN, Jennifer A.
 KALLICK, Deborah A.
 GANDHI, Aameena R.
 ARVIZU, Chandra
 ISON, Craig H.
 TANG, Y. Tom
 AZIMZAI, Yalda
 ELLIOTT, Vicki S.
 SWARNAKAR, Anita
 RAMKUMAR, Jayalaxmi
 NGUYEN, Danniel B.
 TRIBOULEY, Catherine M.
 LO, Terence P.
 AU-YOUNG, Janice
 THANGAVELU, Kavitha
 KEARNEY, Liam

<120> PROTEASES

<130> PI-0263 PCT

<140> To Be Assigned
 <141> Herewith

<150> 60/241,573; 60/243,643; 60/245,256; 60/248,395; 60/249,826
 60/252,303; 60/250,981

<151> 2000-10-18; 2000-10-25; 2000-11-02; 2000-11-13; 2000-11-16
 2000-11-20; 2000-12-01

<160> 32

<170> PERL Program

<210> 1
 <211> 334
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6926819CD1

<400> 1
 Met Asn Pro Ser Leu Leu Leu Ala Ala Phe Phe Leu Gly Ile Ala
 1 5 10 15
 Ser Ala Ala Leu Thr Arg Asp His Ser Leu Asp Ala Gln Trp Thr
 20 25 30

Lys Trp Lys Ala Lys His Lys Arg Leu Tyr Gly Met Asn Arg Asn
 35 40 45
 His Trp Ile Arg Val Leu Trp Glu Lys Asp Val Lys Met Ile Glu
 50 55 60
 Gln His Asn Gln Glu Tyr Ser Gln Gly Lys His Ser Phe Thr Met
 65 70 75
 Ala Met Asn Ala Phe Gly Asp Met Val Ser Glu Glu Phe Arg Gln
 80 85 90
 Val Met Asn Gly Phe Gln Tyr Gln Lys His Arg Lys Gly Lys Gln
 95 100 105
 Phe Gln Glu Arg Leu Leu Glu Ile Pro Thr Ser Val Asp Trp
 110 115 120
 Arg Glu Lys Gly Tyr Met Thr Pro Val Lys Asp Gln Gln Gly Gln
 125 130 135
 Cys Gly Ser Cys Trp Ala Phe Ser Ala Thr Gly Ala Leu Glu Gly
 140 145 150
 Gln Met Phe Trp Lys Thr Gly Lys Leu Ile Ser Leu Asn Glu Gln
 155 160 165
 Asn Leu Val Asp Cys Ser Gly Pro Gln Gly Asn Glu Gly Cys Asn
 170 175 180
 Gly Asp Phe Met Asp Asn Pro Phe Arg Tyr Val Gln Glu Asn Gly
 185 190 195
 Gly Leu Asp Ser Glu Ala Ser Tyr Pro Tyr Glu Gly Lys Val Lys
 200 205 210
 Thr Cys Arg Tyr Asn Pro Lys Tyr Ser Ala Ala Asn Asp Thr Gly
 215 220 225
 Phe Val Asp Ile Pro Ser Arg Glu Lys Asp Leu Ala Lys Ala Val
 230 235 240
 Ala Thr Val Gly Pro Ile Ser Val Ala Val Gly Ala Ser His Val
 245 250 255
 Phe Phe Gln Phe Tyr Lys Lys Gly Ile Tyr Phe Glu Pro Arg Cys
 260 265 270
 Asp Pro Glu Gly Leu Asp His Ala Met Leu Val Val Gly Tyr Ser
 275 280 285
 Tyr Glu Gly Ala Asp Ser Asp Asn Asn Lys Tyr Trp Leu Val Lys
 290 295 300
 Asn Ser Trp Gly Lys Asn Trp Gly Met Asp Gly Tyr Ile Lys Met
 305 310 315
 Ala Lys Asp Arg Arg Asn Asn Cys Gly Ile Ala Thr Ala Ala Ser
 320 325 330
 Tyr Pro Thr Val

<210> 2

<211> 511

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473526CD1

<400> 2

Met Ser Leu Trp Pro Pro Phe Arg Cys Arg Trp Lys Leu Ala Pro
 1 5 10 15
 Arg Tyr Ser Arg Arg Ala Ser Pro Gln Gln Pro Gln Gln Asp Phe
 20 25 30
 Glu Ala Leu Leu Ala Glu Cys Leu Arg Asn Gly Cys Leu Phe Glu
 35 40 45
 Asp Thr Ser Phe Pro Ala Thr Leu Ser Ser Ile Gly Ser Gly Ser
 50 55 60
 Leu Leu Gln Lys Leu Pro Pro Arg Leu Gln Trp Lys Arg Pro Pro
 65 70 75

Glu	Leu	His	Ser	Asn	Pro	Gln	Phe	Tyr	Phe	Ala	Lys	Ala	Lys	Arg
				80					85					90
Leu	Asp	Leu	Cys	Gln	Gly	Ile	Val	Gly	Asp	Cys	Trp	Phe	Leu	Ala
				95					100					105
Ala	Leu	Gln	Ala	Leu	Ala	Leu	His	Gln	Asp	Ile	Leu	Ser	Arg	Val
				110					115					120
Val	Pro	Leu	Asn	Gln	Ser	Phe	Thr	Glu	Lys	Tyr	Ala	Gly	Ile	Phe
				125					130					135
Arg	Phe	Trp	Phe	Trp	His	Tyr	Gly	Asn	Trp	Val	Pro	Val	Val	Ile
				140					145					150
Asp	Asp	Arg	Leu	Pro	Val	Asn	Glu	Ala	Gly	Gln	Leu	Val	Phe	Val
				155					160					165
Ser	Ser	Thr	Tyr	Lys	Asn	Leu	Phe	Trp	Gly	Ala	Leu	Leu	Glu	Lys
				170					175					180
Ala	Tyr	Ala	Lys	Leu	Ser	Gly	Ser	Tyr	Glu	Asp	Leu	Gln	Ser	Gly
				185					190					195
Gln	Val	Ser	Glu	Ala	Leu	Val	Asp	Phe	Thr	Gly	Gly	Val	Thr	Met
				200					205					210
Thr	Ile	Asn	Leu	Ala	Glu	Ala	His	Gly	Asn	Leu	Trp	Asp	Ile	Leu
				215					220					225
Ile	Glu	Ala	Thr	Tyr	Asn	Arg	Thr	Leu	Ile	Gly	Cys	Gln	Thr	His
				230					235					240
Ser	Gly	Glu	Lys	Ile	Leu	Glu	Asn	Gly	Leu	Val	Glu	Gly	His	Ala
				245					250					255
Tyr	Thr	Leu	Thr	Gly	Ile	Arg	Lys	Val	Thr	Cys	Lys	His	Arg	Pro
				260					265					270
Glu	Tyr	Leu	Val	Lys	Leu	Arg	Asn	Pro	Trp	Gly	Lys	Val	Glu	Trp
				275					280					285
Lys	Gly	Asp	Trp	Ser	Asp	Ser	Ser	Ser	Lys	Trp	Glu	Leu	Leu	Ser
				290					295					300
Pro	Lys	Glu	Lys	Ile	Leu	Leu	Leu	Arg	Lys	Asp	Asn	Asp	Gly	Glu
				305					310					315
Phe	Trp	Met	Thr	Leu	Gln	Asp	Phe	Lys	Thr	His	Phe	Val	Leu	Leu
				320					325					330
Val	Ile	Cys	Lys	Leu	Thr	Pro	Gly	Leu	Leu	Ser	Gln	Glu	Ala	Ala
				335					340					345
Gln	Lys	Trp	Thr	Tyr	Thr	Met	Arg	Glu	Gly	Arg	Trp	Glu	Lys	Arg
				350					355					360
Ser	Thr	Ala	Gly	Gly	Gln	Arg	Gln	Leu	Leu	Gln	Asp	Thr	Phe	Trp
				365					370					375
Lys	Asn	Pro	Gln	Phe	Leu	Leu	Ser	Val	Trp	Arg	Pro	Glu	Glu	Gly
				380					385					390
Arg	Arg	Ser	Leu	Arg	Pro	Cys	Ser	Val	Leu	Val	Ser	Leu	Leu	Gln
				395					400					405
Lys	Pro	Arg	His	Arg	Cys	Arg	Lys	Arg	Lys	Pro	Leu	Leu	Ala	Ile
				410					415					420
Gly	Phe	Tyr	Leu	Tyr	Arg	Met	Asn	Lys	Tyr	His	Asp	Asp	Gln	Arg
				425					430					435
Arg	Leu	Pro	Pro	Glu	Phe	Phe	Gln	Arg	Asn	Thr	Pro	Leu	Ser	Gln
				440					445					450
Pro	Asp	Arg	Phe	Leu	Lys	Glu	Lys	Glu	Val	Ser	Gln	Glu	Leu	Cys
				455					460					465
Leu	Glu	Pro	Gly	Thr	Tyr	Leu	Ile	Val	Pro	Ala	Tyr	Trp	Arg	Pro
				470					475					480
Thr	Arg	Ser	Gln	Ser	Ser	Ser	Ser	Gly	Ser	Ser	Pro	Gly	Ser	Thr
				485					490					495
Ser	Phe	Met	Lys	Leu	Ala	Ala	Ile	Leu	Val	Ser	Ser	Ser	Gln	Arg
				500					505					510
Arg														

<210> 3

<211> 812

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478443CD1

<400> 3

```

Met Gly Trp Arg Pro Arg Arg Ala Arg Gly Thr Pro Leu Leu Leu
 1          5          10          15
Leu Leu Leu Leu Leu Leu Leu Trp Pro Val Pro Gly Ala Gly Val
          20          25          30
Leu Gln Gly His Ile Pro Gly Gln Pro Val Thr Pro His Trp Val
          35          40          45
Leu Asp Gly Gln Pro Trp Arg Thr Val Ser Leu Glu Glu Pro Val
          50          55          60
Ser Lys Pro Asp Met Gly Leu Val Ala Leu Glu Ala Glu Gly Gln
          65          70          75
Glu Leu Leu Leu Glu Leu Glu Lys Asn His Arg Leu Leu Ala Pro
          80          85          90
Gly Tyr Ile Glu Thr His Tyr Gly Pro Asp Gly Gln Pro Val Val
          95          100          105
Leu Ala Pro Asn His Thr Asp His Cys His Tyr Gln Gly Arg Val
          110          115          120
Arg Gly Phe Pro Asp Ser Trp Val Val Leu Cys Thr Cys Ser Gly
          125          130          135
Met Ser Gly Leu Ile Thr Leu Ser Arg Asn Ala Ser Tyr Tyr Leu
          140          145          150
Arg Pro Trp Pro Pro Arg Gly Ser Lys Asp Phe Ser Thr His Glu
          155          160          165
Ile Phe Arg Met Glu Gln Leu Leu Thr Trp Lys Gly Thr Cys Gly
          170          175          180
His Arg Asp Pro Gly Asn Lys Ala Gly Met Thr Ser Leu Pro Gly
          185          190          195
Gly Pro Gln Ser Arg Gly Arg Arg Glu Ala Arg Arg Thr Arg Lys
          200          205          210
Tyr Leu Glu Leu Tyr Ile Val Ala Asp His Thr Leu Phe Leu Thr
          215          220          225
Arg His Arg Asn Leu Asn His Thr Lys Gln Arg Leu Leu Glu Val
          230          235          240
Ala Asn Tyr Val Asp Gln Leu Leu Arg Thr Leu Asp Ile Gln Val
          245          250          255
Ala Leu Thr Gly Leu Glu Val Trp Thr Glu Arg Asp Arg Ser Arg
          260          265          270
Val Thr Gln Asp Ala Asn Ala Thr Leu Trp Ala Phe Leu Gln Trp
          275          280          285
Arg Arg Gly Leu Trp Ala Gln Arg Pro His Asp Ser Ala Gln Leu
          290          295          300
Leu Thr Gly Arg Ala Phe Gln Gly Ala Thr Val Gly Leu Ala Pro
          305          310          315
Val Glu Gly Met Cys Arg Ala Glu Ser Ser Gly Gly Val Ser Thr
          320          325          330
Asp His Ser Glu Leu Pro Ile Gly Ala Ala Ala Thr Met Ala His
          335          340          345
Glu Ile Gly His Ser Leu Gly Leu Ser His Asp Pro Asp Gly Cys
          350          355          360
Cys Val Glu Ala Ala Ala Glu Ser Gly Gly Cys Val Met Ala Ala
          365          370          375
Ala Thr Gly His Pro Phe Pro Arg Val Phe Ser Ala Cys Ser Arg
          380          385          390
Arg Gln Leu Arg Ala Phe Phe Arg Lys Gly Gly Gly Ala Cys Leu
          395          400          405
Ser Asn Ala Pro Asp Pro Gly Leu Pro Val Pro Pro Ala Leu Cys

```

	410		415		420
Gly Asn Gly Phe Val Glu Ala Gly Glu Glu Cys Asp Cys Gly Pro					
	425		430		435
Gly Gln Glu Cys Arg Asp Leu Cys Cys Phe Ala His Asn Cys Ser					
	440		445		450
Leu Arg Pro Gly Ala Gln Cys Ala His Gly Asp Cys Cys Val Arg					
	455		460		465
Cys Leu Leu Lys Pro Ala Gly Ala Leu Cys Arg Gln Ala Met Gly					
	470		475		480
Asp Cys Asp Leu Pro Glu Phe Cys Thr Gly Thr Ser Ser His Cys					
	485		490		495
Pro Pro Asp Val Tyr Leu Leu Asp Gly Ser Pro Cys Ala Arg Gly					
	500		505		510
Ser Gly Tyr Cys Trp Asp Gly Ala Cys Pro Thr Leu Glu Gln Gln					
	515		520		525
Cys Gln Gln Leu Trp Gly Pro Gly Ser His Pro Ala Pro Glu Ala					
	530		535		540
Cys Phe Gln Val Val Asn Ser Ala Gly Asp Ala His Gly Asn Cys					
	545		550		555
Gly Gln Asp Ser Glu Gly His Phe Leu Pro Cys Ala Gly Arg Asp					
	560		565		570
Ala Leu Cys Gly Lys Leu Gln Cys Gln Gly Gly Lys Pro Ser Leu					
	575		580		585
Leu Ala Pro His Met Val Pro Val Asp Ser Thr Val His Leu Asp					
	590		595		600
Gly Gln Glu Val Thr Cys Arg Gly Ala Leu Ala Leu Pro Ser Ala					
	605		610		615
Gln Leu Asp Leu Leu Gly Leu Gly Leu Val Glu Pro Gly Thr Gln					
	620		625		630
Cys Gly Pro Arg Met Val Cys Gln Ser Arg Arg Cys Arg Lys Asn					
	635		640		645
Ala Phe Gln Glu Leu Gln Arg Cys Leu Thr Ala Cys His Ser His					
	650		655		660
Gly Val Cys Asn Ser Asn His Asn Cys His Cys Ala Pro Gly Trp					
	665		670		675
Ala Pro Pro Phe Cys Asp Lys Pro Gly Phe Gly Gly Ser Met Asp					
	680		685		690
Ser Gly Pro Val Gln Ala Glu Asn His Asp Thr Phe Leu Leu Ala					
	695		700		705
Met Leu Leu Ser Val Leu Leu Pro Leu Leu Pro Gly Ala Gly Leu					
	710		715		720
Ala Trp Cys Cys Tyr Arg Leu Pro Gly Ala His Leu Gln Arg Cys					
	725		730		735
Ser Trp Gly Cys Arg Arg Asp Pro Ala Cys Ser Gly Pro Lys Asp					
	740		745		750
Gly Pro His Arg Asp His Pro Leu Gly Gly Val His Pro Met Glu					
	755		760		765
Leu Gly Pro Thr Ala Thr Gly Gln Pro Trp Pro Leu Asp Pro Glu					
	770		775		780
Asn Ser His Glu Pro Ser Ser His Pro Glu Lys Pro Leu Pro Ala					
	785		790		795
Val Ser Pro Asp Pro Gln Asp Gln Val Gln Met Pro Arg Ser Cys					
	800		805		810
Leu Trp					

<210> 4

<211> 1236

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<400> 4

6/39

Asn	Phe	Gly	Met	Ile	His	Asp	Gly	Glu	Gly	Asn	Met	Cys	Lys	Lys
				455					460					465
Ser	Glu	Gly	Asn	Ile	Met	Ser	Pro	Thr	Leu	Ala	Gly	Arg	Asn	Gly
				470					475					480
Val	Phe	Ser	Trp	Ser	Pro	Cys	Ser	Arg	Gln	Tyr	Leu	His	Lys	Phe
				485					490					495
Leu	Ser	Thr	Ala	Gln	Ala	Ile	Cys	Leu	Ala	Asp	Gln	Pro	Lys	Pro
				500					505					510
Val	Lys	Glu	Tyr	Lys	Tyr	Pro	Glu	Lys	Leu	Pro	Gly	Glu	Leu	Tyr
				515					520					525
Asp	Ala	Asn	Thr	Gln	Cys	Lys	Trp	Gln	Phe	Gly	Glu	Lys	Ala	Lys
				530					535					540
Leu	Cys	Met	Leu	Asp	Phe	Lys	Lys	Asp	Ile	Cys	Lys	Ala	Leu	Trp
				545					550					555
Cys	His	Arg	Ile	Gly	Arg	Lys	Cys	Glu	Thr	Lys	Phe	Met	Pro	Ala
				560					565					570
Ala	Glu	Gly	Thr	Ile	Cys	Gly	His	Asp	Met	Trp	Cys	Arg	Gly	Gly
				575					580					585
Gln	Cys	Val	Lys	Tyr	Gly	Asp	Glu	Gly	Pro	Lys	Pro	Thr	His	Gly
				590					595					600
His	Trp	Ser	Asp	Trp	Ser	Ser	Trp	Ser	Pro	Cys	Ser	Arg	Thr	Cys
				605					610					615
Gly	Gly	Gly	Val	Ser	His	Arg	Ser	Arg	Leu	Cys	Thr	Asn	Pro	Lys
				620					625					630
Pro	Ser	His	Gly	Gly	Lys	Phe	Cys	Glu	Gly	Ser	Thr	Arg	Thr	Leu
				635					640					645
Lys	Leu	Cys	Asn	Ser	Gln	Lys	Cys	Pro	Arg	Asp	Ser	Val	Asp	Phe
				650					655					660
Arg	Ala	Ala	Gln	Cys	Ala	Glu	His	Asn	Ser	Arg	Arg	Phe	Arg	Gly
				665					670					675
Arg	His	Tyr	Lys	Trp	Lys	Pro	Tyr	Thr	Gln	Val	Glu	Asp	Gln	Asp
				680					685					690
Leu	Cys	Lys	Leu	Tyr	Cys	Ile	Ala	Glu	Gly	Phe	Asp	Phe	Phe	Phe
				695					700					705
Ser	Leu	Ser	Asn	Lys	Val	Lys	Asp	Gly	Thr	Pro	Cys	Ser	Glu	Asp
				710					715					720
Ser	Arg	Asn	Val	Cys	Ile	Asp	Gly	Ile	Cys	Glu	Arg	Val	Gly	Cys
				725					730					735
Asp	Asn	Val	Leu	Gly	Ser	Asp	Ala	Val	Glu	Asp	Val	Cys	Gly	Val
				740					745					750
Cys	Asn	Gly	Asn	Asn	Ser	Ala	Cys	Thr	Ile	His	Arg	Gly	Leu	Tyr
				755					760					765
Thr	Lys	His	His	His	Thr	Asn	Gln	Tyr	Tyr	His	Met	Val	Thr	Ile
				770					775					780
Pro	Ser	Gly	Ala	Arg	Ser	Ile	Arg	Ile	Tyr	Glu	Met	Asn	Val	Ser
				785					790					795
Thr	Ser	Tyr	Ile	Ser	Val	Arg	Asn	Ala	Leu	Arg	Arg	Tyr	Tyr	Leu
				800					805					810
Asn	Gly	His	Trp	Thr	Val	Asp	Trp	Pro	Gly	Arg	Tyr	Lys	Phe	Ser
				815					820					825
Gly	Thr	Thr	Phe	Asp	Tyr	Arg	Arg	Ser	Tyr	Asn	Glu	Pro	Glu	Asn
				830					835					840
Leu	Ile	Ala	Thr	Gly	Pro	Thr	Asn	Glu	Thr	Leu	Ile	Val	Glu	Leu
				845					850					855
Leu	Phe	Gln	Gly	Arg	Asn	Pro	Gly	Val	Ala	Trp	Glu	Tyr	Ser	Met
				860					865					870
Pro	Arg	Leu	Gly	Thr	Glu	Lys	Gln	Pro	Pro	Ala	Gln	Pro	Ser	Tyr
				875					880					885
Thr	Trp	Ala	Ile	Val	Arg	Ser	Glu	Cys	Ser	Val	Ser	Cys	Gly	Gly
				890					895					900
Gly	Gln	Met	Thr	Val	Arg	Glu	Gly	Cys	Tyr	Arg	Asp	Leu	Lys	Phe
				905					910					915
Gln	Val	Asn	Met	Ser	Phe	Cys	Asn	Pro	Lys	Thr	Arg	Pro	Val	Thr

```

          920          925          930
Gly Leu Val Pro Cys Lys Val Ser Ala Cys Pro Pro Ser Trp Ser
          935          940          945
Val Gly Asn Trp Ser Ala Cys Ser Arg Thr Cys Gly Gly Gly Ala
          950          955          960
Gln Ser Arg Pro Val Gln Cys Thr Arg Arg Val His Tyr Asp Ser
          965          970          975
Glu Pro Val Pro Ala Gly Leu Cys Pro Gln Leu Val Pro Pro Ala
          980          985          990
Gly Arg Pro Ala Thr Leu Arg Ala Ala His Leu His Gly Ala Pro
          995          1000          1005
Gly Pro Gly Gln Ser Ala His Thr Pro Val Gly Arg Val Glu Glu
          1010          1015          1020
Arg Ala Val Ala Cys Lys Ser Thr Asn Pro Ser Ala Arg Ala Gln
          1025          1030          1035
Leu Leu Pro Asp Ala Val Cys Thr Ser Glu Pro Lys Pro Arg Met
          1040          1045          1050
His Glu Ala Cys Leu Leu Gln Arg Cys His Lys Pro Lys Lys Leu
          1055          1060          1065
Gln Trp Leu Val Ser Ala Trp Ser Gln Cys Ser Val Thr Cys Glu
          1070          1075          1080
Arg Gly Thr Gln Lys Arg Phe Leu Lys Cys Ala Glu Lys Tyr Val
          1085          1090          1095
Ser Gly Lys Tyr Arg Glu Leu Ala Ser Lys Lys Cys Ser His Leu
          1100          1105          1110
Pro Lys Pro Ser Leu Glu Leu Glu Arg Ala Cys Ala Pro Leu Pro
          1115          1120          1125
Cys Pro Arg His Pro Pro Phe Ala Ala Ala Gly Pro Ser Arg Gly
          1130          1135          1140
Ser Trp Phe Ala Ser Pro Trp Ser Gln Cys Thr Ala Ser Cys Gly
          1145          1150          1155
Gly Gly Val Gln Thr Arg Ser Val Gln Cys Leu Ala Gly Gly Arg
          1160          1165          1170
Pro Ala Ser Gly Cys Leu Leu His Gln Lys Pro Ser Ala Ser Leu
          1175          1180          1185
Ala Cys Asn Thr His Phe Cys Pro Ile Ala Glu Lys Lys Asp Ala
          1190          1195          1200
Phe Cys Lys Asp Tyr Phe His Trp Cys Tyr Leu Val Pro Gln His
          1205          1210          1215
Gly Met Cys Ser His Lys Phe Tyr Gly Lys Gln Cys Cys Lys Thr
          1220          1225          1230
Cys Ser Lys Ser Asn Leu
          1235

```

<210> 5

<211> 304

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483438CD1

<400> 5

```

Met Gly Leu Arg Ala Gly Pro Ile Leu Leu Leu Leu Trp Leu
  1          5          10          15
Leu Pro Gly Ala His Trp Asp Val Leu Pro Ser Glu Cys Gly His
          20          25          30
Ser Lys Glu Ala Gly Arg Ile Val Gly Gly Gln Asp Thr Gln Glu
          35          40          45
Gly Arg Trp Pro Trp Gln Val Gly Leu Trp Leu Thr Ser Val Gly
          50          55          60
His Val Cys Gly Gly Ser Leu Ile His Pro Arg Trp Val Leu Thr

```

	65		70		75
Ala Ala His Cys Phe	Leu Arg Ser Glu Asp	Pro Gly Leu Tyr His			
	80		85		90
Val Lys Val Gly Gly	Leu Thr Pro Ser Leu	Ser Glu Pro His Ser			
	95		100		105
Ala Leu Val Ala Val	Arg Arg Leu Leu Val	His Ser Ser Tyr His			
	110		115		120
Gly Thr Thr Thr Ser	Gly Asp Ile Ala Leu	Met Glu Leu Asp Ser			
	125		130		135
Pro Leu Gln Ala Ser	Gln Phe Ser Pro Ile	Cys Leu Pro Gly Pro			
	140		145		150
Gln Thr Pro Leu Ala	Ile Gly Thr Val Cys	Trp Val Asn Gly Leu			
	155		160		165
Gly Glu Val Ala Val	Pro Leu Leu Asp Ser	Asn Met Cys Glu Leu			
	170		175		180
Met Tyr His Leu Gly	Glu Pro Ser Leu Ala	Gly Gln Arg Leu Ile			
	185		190		195
Gln Asp Asp Met Leu	Cys Ala Gly Ser Val	Gln Gly Lys Lys Asp			
	200		205		210
Ser Cys Gln Gly Asp	Ser Gly Gly Pro Leu	Val Cys Pro Ile Asn			
	215		220		225
Asp Thr Trp Ile Gln	Ala Gly Ile Val Ser	Trp Gly Phe Gly Cys			
	230		235		240
Ala Arg Pro Phe Arg	Pro Gly Val Tyr Thr	Gln Val Leu Ser Tyr			
	245		250		255
Thr Asp Trp Ile Gln	Arg Thr Leu Ala Glu	Ser His Ser Gly Met			
	260		265		270
Ser Gly Ala Arg Pro	Gly Ala Pro Gly Ser	His Ser Gly Thr Ser			
	275		280		285
Arg Ser His Pro Val	Leu Leu Leu Glu Leu	Leu Thr Val Cys Leu			
	290		295		300
Leu Gly Ser Leu					

<210> 6

<211> 980

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7246467CD1

<400> 6

Met Ser Pro Leu Lys	Ile His Gly Pro Ile	Arg Ile Arg Ser Met
1	5	10
Gln Thr Gly Ile Thr	Lys Trp Lys Glu Gly	Ser Phe Glu Ile Val
	20	25
Glu Lys Glu Asn Lys	Val Ser Leu Val Val	His Tyr Asn Thr Gly
	35	40
Gly Ile Pro Arg Ile	Phe Gln Leu Ser His	Asn Ile Lys Asn Val
	50	55
Val Leu Arg Pro Ser	Gly Ala Lys Gln Ser	Arg Leu Met Leu Thr
	65	70
Leu Gln Asp Asn Ser	Phe Leu Ser Ile Asp	Lys Val Pro Ser Lys
	80	85
Asp Ala Glu Glu Met	Arg Leu Phe Leu Asp	Ala Val His Gln Asn
	95	100
Arg Leu Pro Ala Ala	Met Lys Pro Ser Gln	Gly Ser Gly Ser Phe
	110	115
Gly Ala Ile Leu Gly	Ser Arg Thr Ser Gln	Lys Glu Thr Ser Arg
	125	130
Gln Leu Ser Tyr Ser	Asp Asn Gln Ala Ser	Ala Lys Arg Gly Ser

	140		145		150
Leu Glu Thr Lys Asp	Asp Ile Pro Phe Arg	Lys Val Leu Gly Asn			
	155		160		165
Pro Gly Arg Gly Ser	Ile Lys Thr Val Ala	Gly Ser Gly Ile Ala			
	170		175		180
Arg Thr Ile Pro Ser	Leu Thr Ser Thr Ser	Thr Pro Leu Arg Ser			
	185		190		195
Gly Leu Leu Glu Asn	Arg Thr Glu Lys Arg	Lys Arg Met Ile Ser			
	200		205		210
Thr Gly Ser Glu Leu	Asn Glu Asp Tyr Pro	Lys Glu Asn Asp Ser			
	215		220		225
Ser Ser Asn Asn Lys	Ala Met Thr Asp Pro	Ser Arg Lys Tyr Leu			
	230		235		240
Thr Ser Ser Arg Glu	Lys Gln Leu Ser Leu	Lys Gln Ser Glu Glu			
	245		250		255
Asn Arg Thr Ser Gly	Gly Leu Leu Pro Leu	Gln Ser Ser Ser Phe			
	260		265		270
Tyr Gly Ser Arg Ala	Gly Ser Lys Glu His	Ser Ser Gly Gly Thr			
	275		280		285
Asn Leu Asp Arg Thr	Asn Val Ser Ser Gln	Thr Pro Ser Ala Lys			
	290		295		300
Arg Ser Leu Gly Phe	Leu Pro Gln Pro Val	Pro Leu Ser Val Lys			
	305		310		315
Lys Leu Arg Cys Asn	Gln Asp Tyr Thr Gly	Trp Asn Lys Pro Arg			
	320		325		330
Val Pro Leu Ser Ser	His Gln Gln Gln Gln	Leu Gln Gly Phe Ser			
	335		340		345
Asn Leu Gly Asn Thr	Cys Tyr Met Asn Ala	Ile Leu Gln Ser Leu			
	350		355		360
Phe Ser Leu Gln Ser	Phe Ala Asn Asp Leu	Leu Lys Gln Gly Ile			
	365		370		375
Pro Trp Lys Lys Ile	Pro Leu Asn Ala Leu	Ile Arg Arg Phe Ala			
	380		385		390
His Leu Leu Val Lys	Lys Asp Ile Cys Asn	Ser Glu Thr Lys Lys			
	395		400		405
Asp Leu Leu Lys Lys	Val Lys Asn Ala Ile	Ser Ala Thr Ala Glu			
	410		415		420
Arg Phe Ser Gly Tyr	Met Gln Asn Asp Ala	His Glu Phe Leu Ser			
	425		430		435
Gln Cys Leu Asp Gln	Leu Lys Glu Asp Met	Glu Lys Leu Asn Lys			
	440		445		450
Thr Trp Lys Thr Glu	Pro Val Ser Gly Glu	Glu Asn Ser Pro Asp			
	455		460		465
Ile Ser Ala Thr Arg	Ala Tyr Thr Cys Pro	Val Ile Thr Asn Leu			
	470		475		480
Glu Phe Glu Val Gln	His Ser Ile Ile Cys	Lys Ala Cys Gly Glu			
	485		490		495
Ile Ile Pro Lys Arg	Glu Gln Phe Asn Asp	Leu Ser Ile Asp Leu			
	500		505		510
Pro Arg Arg Lys Lys	Pro Leu Pro Pro Arg	Ser Ile Gln Asp Ser			
	515		520		525
Leu Asp Leu Phe Phe	Arg Ala Glu Glu Leu	Glu Tyr Ser Cys Glu			
	530		535		540
Lys Cys Gly Gly Lys	Cys Ala Leu Val Arg	His Lys Phe Asn Arg			
	545		550		555
Leu Pro Arg Val Leu	Ile Leu His Leu Lys	Arg Tyr Ser Phe Asn			
	560		565		570
Val Ala Leu Ser Leu	Asn Asn Lys Ile Gly	Gln Gln Val Ile Ile			
	575		580		585
Pro Arg Tyr Leu Thr	Leu Ser Ser His Cys	Thr Glu Asn Thr Lys			
	590		595		600
Pro Pro Phe Thr Leu	Gly Trp Ser Ala His	Met Ala Met Ser Arg			
	605		610		615

Pro Leu Lys Ala Ser Gln Met Val Asn Ser Cys Ile Thr Ser Pro
 620 625 630
 Ser Thr Pro Ser Lys Lys Phe Thr Phe Lys Ser Lys Ser Ser Leu
 635 640 645
 Ala Leu Cys Leu Asp Ser Asp Ser Glu Asp Glu Leu Lys Arg Ser
 650 655 660
 Val Ala Leu Ser Gln Arg Leu Cys Glu Met Leu Gly Asn Glu Gln
 665 670 675
 Gln Gln Glu Asp Leu Glu Lys Asp Ser Lys Leu Cys Pro Ile Glu
 680 685 690
 Pro Asp Lys Ser Glu Leu Glu Asn Ser Gly Phe Asp Arg Met Ser
 695 700 705
 Glu Glu Glu Leu Leu Ala Ala Val Leu Glu Ile Ser Lys Arg Asp
 710 715 720
 Ala Ser Pro Ser Leu Ser His Glu Asp Asp Asp Lys Pro Thr Ser
 725 730 735
 Ser Pro Asp Thr Gly Phe Ala Glu Asp Asp Ile Gln Glu Met Pro
 740 745 750
 Glu Asn Pro Asp Thr Met Glu Thr Glu Lys Pro Lys Thr Ile Thr
 755 760 765
 Glu Leu Asp Pro Ala Ser Phe Thr Glu Ile Thr Lys Asp Cys Asp
 770 775 780
 Glu Asn Lys Glu Asn Lys Thr Pro Glu Gly Ser Gln Gly Glu Val
 785 790 795
 Asp Trp Leu Gln Gln Tyr Asp Met Glu Arg Glu Arg Glu Glu Gln
 800 805 810
 Glu Leu Gln Gln Ala Leu Ala Gln Ser Leu Gln Glu Gln Glu Ala
 815 820 825
 Trp Glu Gln Lys Glu Asp Asp Asp Leu Lys Arg Ala Thr Glu Leu
 830 835 840
 Ser Leu Gln Glu Phe Asn Asn Ser Phe Val Asp Ala Leu Gly Ser
 845 850 855
 Asp Glu Asp Ser Gly Asn Glu Asp Val Phe Asp Met Glu Tyr Thr
 860 865 870
 Glu Ala Glu Ala Glu Glu Leu Lys Arg Asn Ala Glu Thr Gly Asn
 875 880 885
 Leu Pro His Ser Tyr Arg Leu Ile Ser Val Val Ser His Ile Gly
 890 895 900
 Ser Thr Ser Ser Ser Gly His Tyr Ile Ser Asp Val Tyr Asp Ile
 905 910 915
 Lys Lys Gln Ala Trp Phe Thr Tyr Asn Asp Leu Glu Val Ser Lys
 920 925 930
 Ile Gln Glu Ala Ala Val Gln Ser Asp Arg Asp Arg Ser Gly Tyr
 935 940 945
 Ile Phe Phe Tyr Met His Lys Glu Ile Phe Asp Glu Leu Leu Glu
 950 955 960
 Thr Glu Lys Asn Ser Gln Ser Leu Ser Thr Glu Val Gly Lys Thr
 965 970 975
 Thr Arg Gln Ala Ser
 980

<210> 7

<211> 1251

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7997881CD1

<400> 7

Met Thr Ile Val Asp Lys Ala Ser Glu Ser Ser Asp Pro Ser Ala
 1 5 10 15

Tyr	Gln	Asn	Gln	Pro	Gly	Ser	Ser	Glu	Ala	Val	Ser	Pro	Gly	Asp	
				20					25					30	
Met	Asp	Ala	Gly	Ser	Ala	Ser	Trp	Gly	Ala	Val	Ser	Ser	Leu	Asn	
				35					40					45	
Asp	Val	Ser	Asn	His	Thr	Leu	Ser	Leu	Gly	Pro	Val	Pro	Gly	Ala	
				50					55					60	
Val	Val	Tyr	Ser	Ser	Ser	Ser	Val	Pro	Asp	Lys	Ser	Lys	Pro	Ser	
				65					70					75	
Pro	Gln	Lys	Asp	Gln	Ala	Leu	Gly	Asp	Gly	Ile	Ala	Pro	Pro	Gln	
				80					85					90	
Lys	Val	Leu	Phe	Pro	Ser	Glu	Lys	Ile	Cys	Leu	Lys	Trp	Gln	Gln	
				95					100					105	
Thr	His	Arg	Val	Gly	Ala	Gly	Leu	Gln	Asn	Leu	Gly	Asn	Thr	Cys	
				110					115					120	
Phe	Ala	Asn	Ala	Ala	Leu	Gln	Cys	Leu	Thr	Tyr	Thr	Pro	Pro	Leu	
				125					130					135	
Ala	Asn	Tyr	Met	Leu	Ser	His	Glu	His	Ser	Lys	Thr	Cys	His	Ala	
				140					145					150	
Glu	Gly	Phe	Cys	Met	Met	Cys	Thr	Met	Gln	Ala	His	Ile	Thr	Gln	
				155					160					165	
Ala	Leu	Ser	Asn	Pro	Gly	Asp	Val	Ile	Lys	Pro	Met	Phe	Val	Ile	
				170					175					180	
Asn	Glu	Met	Arg	Arg	Ile	Ala	Arg	His	Phe	Arg	Phe	Gly	Asn	Gln	
				185					190					195	
Glu	Asp	Ala	His	Glu	Phe	Leu	Gln	Tyr	Thr	Val	Asp	Ala	Met	Gln	
				200					205					210	
Lys	Ala	Cys	Leu	Asn	Gly	Ser	Asn	Lys	Leu	Asp	Arg	His	Thr	Gln	
				215					220					225	
Ala	Thr	Thr	Leu	Val	Cys	Gln	Ile	Phe	Gly	Gly	Tyr	Leu	Arg	Ser	
				230					235					240	
Arg	Val	Lys	Cys	Leu	Asn	Cys	Lys	Gly	Val	Ser	Asp	Thr	Phe	Asp	
				245					250					255	
Pro	Tyr	Leu	Asp	Ile	Thr	Leu	Glu	Ile	Lys	Ala	Ala	Gln	Ser	Val	
				260					265					270	
Asn	Lys	Ala	Leu	Glu	Gln	Phe	Val	Lys	Pro	Glu	Gln	Leu	Asp	Gly	
				275					280					285	
Glu	Asn	Ser	Tyr	Lys	Cys	Ser	Lys	Cys	Lys	Lys	Met	Val	Pro	Ala	
				290					295					300	
Ser	Lys	Arg	Phe	Thr	Ile	His	Arg	Ser	Ser	Asn	Val	Leu	Thr	Leu	
				305					310					315	
Ser	Leu	Lys	Arg	Phe	Ala	Asn	Phe	Thr	Gly	Gly	Lys	Ile	Ala	Lys	
				320					325					330	
Asp	Val	Lys	Tyr	Pro	Glu	Tyr	Leu	Asp	Ile	Arg	Pro	Tyr	Met	Ser	
				335					340					345	
Gln	Pro	Asn	Gly	Glu	Pro	Ile	Val	Tyr	Val	Leu	Tyr	Ala	Val	Leu	
				350					355					360	
Val	His	Thr	Gly	Phe	Asn	Cys	His	Ala	Gly	His	Tyr	Phe	Cys	Tyr	
				365					370					375	
Ile	Lys	Ala	Ser	Asn	Gly	Leu	Trp	Tyr	Gln	Met	Asn	Asp	Ser	Ile	
				380					385					390	
Val	Ser	Thr	Ser	Asp	Ile	Arg	Ser	Val	Leu	Ser	Gln	Gln	Ala	Tyr	
				395					400					405	
Val	Leu	Phe	Tyr	Ile	Arg	Ser	His	Asp	Val	Lys	Asn	Gly	Gly	Glu	
				410					415					420	
Leu	Thr	His	Pro	Thr	His	Ser	Pro	Gly	Gln	Ser	Ser	Pro	Arg	Pro	
				425					430					435	
Val	Ile	Ser	Gln	Arg	Val	Val	Thr	Asn	Lys	Gln	Ala	Ala	Pro	Gly	
				440					445					450	
Phe	Ile	Gly	Pro	Gln	Leu	Pro	Ser	His	Met	Ile	Lys	Asn	Pro	Pro	
				455					460					465	
His	Leu	Asn	Gly	Thr	Gly	Pro	Leu	Lys	Asp	Thr	Pro	Ser	Ser	Ser	
				470					475					480	
Met	Ser	Ser	Pro	Asn	Gly	Asn	Ser	Ser	Val	Asn	Arg	Ala	Ser	Pro	

	485		490		495
Val Asn Ala Ser	Ala Ser Val Gln Asn	Trp Ser Val Asn Arg Ser			
	500		505		510
Ser Val Ile Pro	Glu His Pro Lys Lys	Gln Lys Ile Thr Ile Ser			
	515		520		525
Ile His Asn Lys	Leu Pro Val Arg Gln	Cys Gln Ser Gln Pro Asn			
	530		535		540
Leu His Ser Asn	Ser Leu Glu Asn Pro	Thr Lys Pro Val Pro Ser			
	545		550		555
Ser Thr Ile Thr	Asn Ser Ala Val Gln	Ser Thr Ser Asn Ala Ser			
	560		565		570
Thr Met Ser Val	Ser Ser Lys Val Thr	Lys Pro Ile Pro Arg Ser			
	575		580		585
Glu Ser Cys Ser	Gln Pro Val Met Asn	Gly Lys Ser Lys Leu Asn			
	590		595		600
Ser Ser Val Leu	Val Pro Tyr Gly Ala	Glu Ser Ser Glu Asp Ser			
	605		610		615
Asp Glu Glu Ser	Lys Gly Leu Gly Lys	Glu Asn Gly Ile Gly Thr			
	620		625		630
Ile Val Ser Ser	His Ser Pro Gly Gln	Asp Ala Glu Asp Glu Glu			
	635		640		645
Ala Thr Pro His	Glu Leu Gln Glu Pro	Met Thr Leu Asn Gly Ala			
	650		655		660
Asn Ser Ala Asp	Ser Asp Ser Asp Pro	Lys Glu Asn Gly Leu Ala			
	665		670		675
Pro Asp Gly Ala	Ser Cys Gln Gly Gln	Pro Ala Leu His Ser Glu			
	680		685		690
Asn Pro Phe Ala	Lys Ala Asn Gly Leu	Pro Gly Lys Leu Met Pro			
	695		700		705
Ala Pro Leu Leu	Ser Leu Pro Glu Asp	Lys Ile Leu Glu Thr Phe			
	710		715		720
Arg Leu Ser Asn	Lys Leu Lys Gly Ser	Thr Asp Glu Met Ser Ala			
	725		730		735
Pro Gly Ala Glu	Arg Gly Pro Pro Glu	Asp Arg Asp Ala Glu Pro			
	740		745		750
Gln Pro Gly Ser	Pro Ala Ala Glu Ser	Leu Glu Glu Pro Asp Ala			
	755		760		765
Ala Ala Gly Leu	Ser Ser Thr Lys Lys	Ala Pro Pro Pro Arg Asp			
	770		775		780
Pro Gly Thr Pro	Ala Thr Lys Glu Gly	Ala Trp Glu Ala Met Ala			
	785		790		795
Val Ala Pro Glu	Glu Pro Pro Pro Ser	Ala Gly Glu Asp Ile Val			
	800		805		810
Gly Asp Thr Ala	Pro Pro Asp Leu Cys	Asp Pro Gly Ser Leu Thr			
	815		820		825
Gly Asp Ala Ser	Pro Leu Ser Gln Asp	Ala Lys Gly Met Ile Ala			
	830		835		840
Glu Gly Pro Arg	Asp Ser Ala Leu Ala	Glu Ala Pro Glu Gly Leu			
	845		850		855
Ser Pro Ala Pro	Pro Ala Arg Ser Glu	Glu Pro Cys Glu Gln Pro			
	860		865		870
Leu Leu Val His	Pro Ser Gly Asp His	Ala Arg Asp Ala Gln Asp			
	875		880		885
Pro Ser Gln Ser	Leu Gly Ala Pro Glu	Ala Ala Glu Arg Pro Pro			
	890		895		900
Ala Pro Val Leu	Asp Met Ala Pro Ala	Gly His Pro Glu Gly Asp			
	905		910		915
Ala Glu Pro Ser	Pro Gly Glu Arg Val	Glu Asp Ala Ala Ala Pro			
	920		925		930
Lys Ala Pro Gly	Pro Ser Pro Ala Lys	Glu Lys Ile Gly Ser Leu			
	935		940		945
Arg Lys Val Asp	Arg Gly His Tyr Arg	Ser Arg Arg Glu Arg Ser			
	950		955		960

Ser Ser Gly Glu Pro Ala Arg Glu Ser Arg Ser Lys Thr Glu Gly
 965 970 975
 His Arg His Arg Arg Arg Arg Thr Cys Pro Arg Glu Arg Asp Arg
 980 985 990
 Gln Asp Arg His Ala Pro Glu His His Pro Gly His Gly Asp Arg
 995 1000 1005
 Leu Ser Pro Gly Glu Arg Arg Ser Leu Gly Arg Cys Ser His His
 1010 1015 1020
 His Ser Arg His Arg Ser Gly Val Glu Leu Asp Trp Val Arg His
 1025 1030 1035
 His Tyr Thr Glu Gly Glu Arg Gly Trp Gly Arg Glu Lys Phe Tyr
 1040 1045 1050
 Pro Asp Arg Pro Arg Trp Asp Arg Cys Arg Tyr Tyr His Asp Arg
 1055 1060 1065
 Tyr Ala Leu Tyr Ala Ala Arg Asp Trp Lys Pro Phe His Gly Gly
 1070 1075 1080
 Arg Glu His Glu Arg Ala Gly Leu His Glu Arg Pro His Lys Asp
 1085 1090 1095
 His Asn Arg Gly Arg Arg Gly Cys Glu Pro Ala Arg Glu Arg Glu
 1100 1105 1110
 Arg His Arg Pro Ser Ser Pro Arg Ala Gly Ala Pro His Ala Leu
 1115 1120 1125
 Ala Pro His Pro Asp Arg Phe Ser His Asp Arg Thr Ala Leu Val
 1130 1135 1140
 Ala Gly Asp Asn Cys Asn Leu Ser Asp Arg Phe His Glu His Glu
 1145 1150 1155
 Asn Gly Lys Ser Arg Lys Arg Arg His Asp Ser Val Glu Asn Ser
 1160 1165 1170
 Asp Ser His Val Glu Lys Lys Ala Arg Arg Ser Glu Gln Lys Asp
 1175 1180 1185
 Pro Leu Glu Glu Pro Lys Ala Lys Lys His Lys Lys Ser Lys Lys
 1190 1195 1200
 Lys Lys Lys Ser Lys Asp Lys His Arg Asp Arg Asp Ser Arg His
 1205 1210 1215
 Gln Gln Asp Ser Asp Leu Ser Ala Ala Cys Ser Asp Ala Asp Leu
 1220 1225 1230
 His Arg His Lys Lys Lys Glu Glu Glu Lys Glu Glu Thr Phe Lys
 1235 1240 1245
 Lys Ile Arg Gly Leu Cys
 1250

<210> 8

<211> 1128

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7484378CD1

<400> 8

Met Glu Pro Thr Val Ala Asp Val His Leu Val Pro Arg Thr Thr
 1 5 10 15
 Lys Glu Val Pro Ala Leu Asp Ala Ala Cys Cys Arg Ala Ala Ser
 20 25 30
 Ile Gly Val Val Ala Thr Ser Leu Val Val Leu Thr Leu Gly Val
 35 40 45
 Leu Leu Gly Gly Met Asn Asn Ser Arg His Ala Ala Leu Arg Ala
 50 55 60
 Ala Thr Leu Pro Gly Lys Val Tyr Ser Val Thr Pro Glu Ala Ser
 65 70 75
 Lys Thr Thr Asn Pro Pro Glu Gly Arg Asn Ser Glu His Ile Arg
 80 85 90

Thr	Ser	Ala	Arg	Thr	Asn	Ser	Gly	His	Thr	Ile	Phe	Lys	Lys	Cys
				95					100					105
Asn	Thr	Gln	Pro	Phe	Leu	Ser	Thr	Gln	Gly	Phe	His	Val	Asp	His
				110					115					120
Thr	Ala	Glu	Leu	Arg	Gly	Ile	Arg	Trp	Thr	Ser	Ser	Leu	Arg	Arg
				125					130					135
Glu	Thr	Ser	Asp	Tyr	His	Arg	Thr	Leu	Thr	Pro	Thr	Leu	Glu	Ala
				140					145					150
Leu	Leu	His	Phe	Leu	Leu	Arg	Pro	Leu	Gln	Thr	Leu	Ser	Leu	Gly
				155					160					165
Leu	Glu	Glu	Glu	Leu	Leu	Gln	Arg	Gly	Ile	Arg	Ala	Arg	Leu	Arg
				170					175					180
Glu	His	Gly	Ile	Ser	Leu	Ala	Ala	Tyr	Gly	Thr	Ile	Val	Ser	Ala
				185					190					195
Glu	Leu	Thr	Gly	Arg	His	Lys	Gly	Pro	Leu	Ala	Glu	Arg	Asp	Phe
				200					205					210
Lys	Ser	Gly	Arg	Cys	Pro	Gly	Asn	Ser	Phe	Ser	Cys	Gly	Asn	Ser
				215					220					225
Gln	Cys	Val	Thr	Lys	Val	Asn	Pro	Glu	Cys	Asp	Asp	Gln	Glu	Asp
				230					235					240
Cys	Ser	Asp	Gly	Ser	Asp	Glu	Ala	His	Cys	Glu	Cys	Gly	Leu	Gln
				245					250					255
Pro	Ala	Trp	Arg	Met	Ala	Gly	Arg	Ile	Val	Gly	Gly	Met	Glu	Ala
				260					265					270
Ser	Pro	Gly	Glu	Phe	Pro	Trp	Gln	Ala	Ser	Leu	Arg	Glu	Asn	Lys
				275					280					285
Glu	His	Phe	Cys	Gly	Ala	Ala	Ile	Ile	Asn	Ala	Arg	Trp	Leu	Val
				290					295					300
Ser	Ala	Ala	His	Cys	Phe	Asn	Glu	Phe	Gln	Asp	Pro	Thr	Lys	Trp
				305					310					315
Val	Ala	Tyr	Val	Gly	Ala	Thr	Tyr	Leu	Ser	Gly	Ser	Glu	Ala	Ser
				320					325					330
Thr	Val	Arg	Ala	Gln	Val	Val	Gln	Ile	Val	Lys	His	Pro	Leu	Tyr
				335					340					345
Asn	Ala	Asp	Thr	Ala	Asp	Phe	Asp	Val	Ala	Val	Leu	Glu	Leu	Thr
				350					355					360
Ser	Pro	Leu	Pro	Phe	Gly	Arg	His	Ile	Gln	Pro	Val	Cys	Leu	Pro
				365					370					375
Ala	Ala	Thr	His	Ile	Phe	Pro	Pro	Ser	Lys	Lys	Cys	Leu	Ile	Ser
				380					385					390
Gly	Trp	Gly	Tyr	Leu	Lys	Glu	Asp	Phe	Arg	Lys	His	Leu	Pro	Arg
				395					400					405
Pro	Ala	Met	Val	Lys	Pro	Glu	Val	Leu	Gln	Lys	Ala	Thr	Val	Glu
				410					415					420
Leu	Leu	Asp	Gln	Ala	Leu	Cys	Ala	Ser	Leu	Tyr	Gly	His	Ser	Leu
				425					430					435
Thr	Asp	Arg	Met	Val	Cys	Ala	Gly	Tyr	Leu	Asp	Gly	Lys	Val	Asp
				440					445					450
Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Glu	Glu	Pro
				455					460					465
Ser	Gly	Arg	Phe	Phe	Leu	Ala	Gly	Ile	Val	Ser	Trp	Gly	Ile	Gly
				470					475					480
Cys	Ala	Glu	Ala	Arg	Arg	Pro	Gly	Val	Tyr	Ala	Arg	Val	Thr	Arg
				485					490					495
Leu	Arg	Asp	Trp	Ile	Leu	Glu	Ala	Thr	Thr	Lys	Ala	Ser	Met	Pro
				500					505					510
Leu	Ala	Pro	Thr	Met	Ala	Pro	Ala	Pro	Ala	Ala	Pro	Ser	Thr	Ala
				515					520					525
Trp	Pro	Thr	Ser	Pro	Glu	Ser	Pro	Val	Val	Ser	Thr	Pro	Thr	Lys
				530					535					540
Ser	Met	Gln	Ala	Leu	Ser	Thr	Val	Pro	Leu	Asp	Trp	Val	Thr	Val
				545					550					555
Pro	Lys	Leu	Gln	Glu	Cys	Gly	Ala	Arg	Pro	Ala	Met	Glu	Lys	Pro

	560		565		570
Thr Arg Val Val	Gly Gly Phe Gly Ala	Ala Ser Gly Glu Val	Pro		
	575		580		585
Trp Gln Val Ser	Leu Lys Glu Gly Ser	Arg His Phe Cys Gly	Ala		
	590		595		600
Thr Val Val Gly	Asp Arg Trp Leu Leu	Ser Ala Ala His Cys	Phe		
	605		610		615
Asn His Thr Lys	Val Glu Gln Val Arg	Ala His Leu Gly Thr	Ala		
	620		625		630
Ser Leu Leu Gly	Leu Gly Gly Ser Pro	Val Lys Ile Gly Leu	Arg		
	635		640		645
Arg Val Val Leu	His Pro Leu Tyr Asn	Pro Gly Ile Leu Asp	Phe		
	650		655		660
Asp Leu Ala Val	Leu Glu Leu Ala Ser	Pro Leu Ala Phe Asn	Lys		
	665		670		675
Tyr Ile Gln Pro	Val Cys Leu Pro Leu	Ala Ile Gln Lys Phe	Pro		
	680		685		690
Val Gly Arg Lys	Cys Met Ile Ser Gly	Trp Gly Asn Thr Gln	Glu		
	695		700		705
Gly Asn Ala Thr	Lys Pro Glu Leu Leu	Gln Lys Ala Ser Val	Gly		
	710		715		720
Ile Ile Asp Gln	Lys Thr Cys Ser Val	Leu Tyr Asn Phe Ser	Leu		
	725		730		735
Thr Asp Arg Met	Ile Cys Ala Gly Phe	Leu Glu Gly Lys Val	Asp		
	740		745		750
Ser Cys Gln Gly	Asp Ser Gly Gly Pro	Leu Ala Cys Glu Glu	Ala		
	755		760		765
Pro Gly Val Phe	Tyr Leu Ala Gly Ile	Val Ser Trp Gly Ile	Gly		
	770		775		780
Cys Ala Gln Val	Lys Lys Pro Gly Val	Tyr Thr Arg Ile Thr	Arg		
	785		790		795
Leu Lys Gly Trp	Ile Leu Glu Ile Met	Ser Ser Gln Pro Leu	Pro		
	800		805		810
Met Ser Pro Pro	Ser Thr Thr Arg Met	Leu Ala Thr Thr Ser	Pro		
	815		820		825
Arg Thr Thr Ala	Gly Leu Thr Val Pro	Gly Ala Thr Pro Ser	Arg		
	830		835		840
Pro Thr Pro Gly	Ala Ala Ser Arg Val	Thr Gly Gln Pro Ala	Asn		
	845		850		855
Ser Thr Leu Ser	Ala Val Ser Thr Thr	Ala Arg Gly Gln Thr	Pro		
	860		865		870
Phe Pro Asp Ala	Pro Glu Ala Thr Thr	His Thr Gln Leu Pro	Asp		
	875		880		885
Cys Gly Leu Ala	Pro Ala Ala Leu Thr	Arg Ile Val Gly Gly	Ser		
	890		895		900
Ala Ala Gly Arg	Gly Glu Trp Pro Trp	Gln Val Ser Leu Trp	Leu		
	905		910		915
Arg Arg Arg Glu	His Arg Cys Gly Ala	Val Leu Val Ala Glu	Arg		
	920		925		930
Trp Leu Leu Ser	Ala Ala His Cys Phe	Asp Val Tyr Gly Asp	Pro		
	935		940		945
Lys Gln Trp Ala	Ala Phe Leu Gly Thr	Pro Phe Leu Ser Gly	Ala		
	950		955		960
Glu Gly Gln Leu	Glu Arg Val Ala Arg	Ile Tyr Lys His Pro	Phe		
	965		970		975
Tyr Asn Leu Tyr	Thr Leu Asp Tyr Asp	Val Ala Leu Leu Glu	Leu		
	980		985		990
Ala Gly Pro Val	Arg Arg Ser Arg Leu	Val Arg Pro Ile Cys	Leu		
	995		1000		1005
Pro Glu Pro Ala	Pro Arg Pro Pro Asp	Gly Thr Arg Cys Val	Ile		
	1010		1015		1020
Thr Gly Trp Gly	Ser Val Arg Glu Gly	Gly Ser Met Ala Arg	Gln		
	1025		1030		1035

```

Leu Gln Lys Ala Ala Val Arg Leu Leu Ser Glu Gln Thr Cys Arg
      1040      1045      1050
Arg Phe Tyr Pro Val Gln Ile Ser Ser Arg Met Leu Cys Ala Gly
      1055      1060      1065
Phe Pro Gln Gly Gly Val Asp Ser Cys Ser Gly Asp Ala Gly Gly
      1070      1075      1080
Pro Leu Ala Cys Arg Glu Pro Ser Gly Arg Trp Val Leu Thr Gly
      1085      1090      1095
Val Thr Ser Trp Gly Tyr Gly Cys Gly Arg Pro His Phe Pro Gly
      1100      1105      1110
Val Tyr Thr Arg Val Ala Ala Val Arg Gly Trp Ile Gly Gln His
      1115      1120      1125
Ile Gln Glu

```

```

<210> 9
<211> 462
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7473143CD1

```

```

<400> 9
Met Ile Pro Phe Thr Glu Leu Gly Gly Arg Gln Gln Lys Arg Arg
  1      5      10      15
Glu Trp Val Gly Gly His Arg Glu His Pro Lys Gly Val Met Gly
      20      25      30
Leu Ala His Arg Gly Met Ala Gly Leu Asp His Asp Val Val Ser
      35      40      45
Asn Gln Cys Thr Ser Gly Lys Ser Pro Lys Ser Glu Arg Gly Ala
      50      55      60
Glu Ala Leu Ala Arg Arg Leu Lys Gly Gly Arg Glu Arg Ala Gly
      65      70      75
Ala Gly Lys Glu Tyr Gly Ile Val Gly Gly Ser Ser Gly His Cys
      80      85      90
Cys Ser Lys Cys Gly Pro Thr Glu Gly Ile Ile Thr Ser Pro Gly
      95      100      105
Ser Met Val Gly Arg Gln Ser Leu Gln Leu His Pro Gly Val Asp
      110      115      120
Leu Asn Leu His Leu Arg Gln Ile Pro Gln Val Met Arg Val His
      125      130      135
Ser Gln Asn Cys Thr Phe Gln Leu His Gly Pro Asn Gly Thr Val
      140      145      150
Glu Ser Pro Gly Phe Pro Tyr Gly Tyr Pro Asn Tyr Ala Asn Cys
      155      160      165
Thr Trp Thr Ile Thr Ala Glu Glu Gln His Arg Ile Gln Leu Val
      170      175      180
Phe Gln Ser Phe Ala Leu Glu Glu Asp Phe Asp Val Leu Ser Val
      185      190      195
Phe Asp Gly Pro Pro Gln Pro Glu Asn Leu Arg Thr Arg Leu Thr
      200      205      210
Gly Phe Gln Leu Pro Ala Thr Ile Val Ser Ala Ala Thr Thr Leu
      215      220      225
Ser Leu Arg Leu Ile Ser Asp Tyr Ala Val Ser Ala Gln Gly Phe
      230      235      240
His Ala Thr Tyr Glu Val Leu Pro Ser His Thr Cys Gly Asn Pro
      245      250      255
Gly Arg Leu Pro Asn Gly Ile Gln Gln Gly Ser Thr Phe Asn Leu
      260      265      270
Gly Asp Lys Val Arg Tyr Ser Cys Asn Leu Gly Phe Phe Leu Glu
      275      280      285

```

Gly	His	Ala	Val	Leu	Thr	Cys	His	Ala	Gly	Ser	Glu	Asn	Ser	Ala	
				290					295					300	
Thr	Trp	Asp	Phe	Pro	Leu	Pro	Ser	Cys	Arg	Ala	Asp	Asp	Ala	Cys	
				305					310					315	
Gly	Gly	Thr	Leu	Arg	Gly	Gln	Ser	Gly	Ile	Ile	Ser	Ser	Pro	His	
				320					325					330	
Phe	Pro	Ser	Glu	Tyr	His	Asn	Asn	Ala	Asp	Cys	Thr	Trp	Thr	Ile	
				335					340					345	
Leu	Ala	Glu	Leu	Gly	Asp	Thr	Ile	Ala	Leu	Val	Phe	Ile	Asp	Phe	
				350					355					360	
Gln	Leu	Glu	Asp	Gly	Tyr	Asp	Phe	Leu	Glu	Val	Thr	Gly	Thr	Glu	
				365					370					375	
Gly	Ser	Ser	Leu	Trp	Phe	Thr	Gly	Ala	Ser	Leu	Pro	Ala	Pro	Val	
				380					385					390	
Ile	Ser	Ser	Lys	Asn	Trp	Leu	Arg	Leu	His	Phe	Thr	Ser	Asp	Gly	
				395					400					405	
Asn	His	Arg	Gln	Arg	Gly	Phe	Ser	Ala	Gln	Tyr	Gln	Val	Lys	Lys	
				410					415					420	
Gln	Ile	Glu	Leu	Lys	Ser	Arg	Gly	Val	Lys	Leu	Met	Pro	Ser	Lys	
				425					430					435	
Asp	Asn	Ser	Gln	Lys	Thr	Ser	Val	Cys	Phe	His	Leu	Thr	Pro	Arg	
				440					445					450	
Ala	Cys	Leu	Ser	Leu	Ser	Ser	Leu	Leu	Pro	Cys	Val				
				455					460						

<210> 10

<211> 659

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4382838CD1

<400> 10

Met	Leu	Trp	Ser	Glu	Arg	Val	Arg	Pro	Ser	Tyr	Ser	Cys	Ile	Ala	
1				5					10					15	
Asn	Asn	Asn	Val	Gly	Asn	Pro	Ala	Lys	Lys	Ser	Thr	Asn	Ile	Ile	
				20					25					30	
Val	Arg	Ala	Leu	Lys	Lys	Gly	Arg	Phe	Trp	Ile	Thr	Pro	Asp	Pro	
				35					40					45	
Tyr	His	Lys	Asp	Asp	Asn	Ile	Gln	Ile	Gly	Arg	Glu	Val	Lys	Ile	
				50					55					60	
Ser	Cys	Gln	Val	Glu	Ala	Val	Pro	Ser	Glu	Glu	Val	Thr	Phe	Ser	
				65					70					75	
Trp	Phe	Lys	Asn	Gly	Arg	Pro	Leu	Arg	Ser	Ser	Glu	Arg	Met	Val	
				80					85					90	
Ile	Thr	Gln	Thr	Asp	Pro	Asp	Val	Ser	Pro	Gly	Thr	Thr	Asn	Leu	
				95					100					105	
Asp	Ile	Ile	Asp	Leu	Lys	Phe	Thr	Asp	Phe	Gly	Thr	Tyr	Thr	Cys	
				110					115					120	
Val	Ala	Ser	Leu	Lys	Gly	Gly	Gly	Ile	Ser	Asp	Ile	Ser	Ile	Asp	
				125					130					135	
Val	Asn	Ile	Ser	Ser	Ser	Thr	Val	Pro	Pro	Asn	Leu	Thr	Val	Pro	
				140					145					150	
Gln	Glu	Lys	Ser	Pro	Leu	Val	Thr	Arg	Glu	Gly	Asp	Thr	Ile	Glu	
				155					160					165	
Leu	Gln	Cys	Gln	Val	Thr	Gly	Lys	Pro	Lys	Pro	Ile	Ile	Leu	Trp	
				170					175					180	
Ser	Arg	Ala	Asp	Lys	Glu	Val	Ala	Met	Pro	Asp	Gly	Ser	Met	Gln	
				185					190					195	
Met	Glu	Ser	Tyr	Asp	Gly	Thr	Leu	Arg	Ile	Val	Asn	Val	Ser	Arg	
				200					205					210	

Glu Met Ser Gly	Met Tyr Arg Cys Gln Thr Ser Gln Tyr Asn Gly	215	220	225
Phe Asn Val Lys	Pro Arg Glu Ala Leu Val Gln Leu Ile Val Gln	230	235	240
Tyr Pro Pro Ala	Val Glu Pro Ala Phe Leu Glu Ile Arg Gln Gly	245	250	255
Gln Asp Arg Ser	Val Thr Met Ser Cys Arg Val Leu Arg Ala Tyr	260	265	270
Pro Ile Arg Val	Leu Thr Tyr Glu Trp Arg Leu Gly Asn Lys Leu	275	280	285
Leu Arg Thr Gly	Gln Phe Asp Ser Gln Glu Tyr Thr Glu Tyr Ala	290	295	300
Val Lys Ser Leu	Ser Asn Glu Asn Tyr Gly Val Tyr Asn Cys Ser	305	310	315
Ile Ile Asn Glu	Ala Gly Ala Gly Arg Cys Ser Phe Leu Val Thr	320	325	330
Gly Lys Ala Tyr	Ala Pro Glu Phe Tyr Tyr Asp Thr Tyr Asn Pro	335	340	345
Val Trp Gln Asn	Arg His Arg Val Tyr Ser Tyr Ser Leu Gln Trp	350	355	360
Thr Gln Met Asn	Pro Asp Ala Val Asp Arg Ile Val Ala Tyr Arg	365	370	375
Leu Gly Ile Arg	Gln Ala Gly Gln Gln Arg Trp Trp Glu Gln Glu	380	385	390
Ile Lys Ile Asn	Gly Asn Ile Gln Lys Gly Glu Leu Ile Thr Tyr	395	400	405
Asn Leu Thr Glu	Leu Ile Lys Pro Glu Ala Tyr Glu Val Arg Leu	410	415	420
Thr Pro Leu Thr	Lys Phe Gly Glu Gly Asp Ser Thr Ile Arg Val	425	430	435
Ile Lys Tyr Ser	Ala Pro Val Asn Pro His Leu Arg Glu Phe His	440	445	450
Arg Gly Phe Glu	Asp Gly Asn Ile Cys Leu Phe Thr Gln Asp Asp	455	460	465
Thr Asp Asn Phe	Asp Trp Thr Lys Gln Ser Thr Ala Thr Arg Asn	470	475	480
Thr Lys Tyr Thr	Pro Asn Thr Gly Pro Asn Ala Asp Arg Ser Gly	485	490	495
Ser Lys Glu Gly	Phe Tyr Met Tyr Ile Glu Thr Ser Arg Pro Arg	500	505	510
Leu Glu Gly Glu	Lys Ala Arg Leu Pro Ser Pro Val Phe Ser Ile	515	520	525
Ala Pro Lys Asn	Pro Tyr Gly Pro Thr Asn Thr Ala Tyr Cys Phe	530	535	540
Ser Phe Phe Tyr	His Met Tyr Gly Gln His Ile Gly Val Leu Asn	545	550	555
Val Tyr Leu Arg	Leu Lys Gly Gln Thr Thr Ile Glu Asn Pro Leu	560	565	570
Trp Ser Ser Ser	Gly Asn Lys Gly Gln Arg Trp Asn Glu Ala His	575	580	585
Val Asn Ile Tyr	Pro Ile Thr Ser Phe Gln Leu Ile Phe Glu Gly	590	595	600
Ile Arg Gly Pro	Gly Ile Glu Gly Asp Ile Ala Ile Asp Asp Val	605	610	615
Ser Ile Ala Glu	Gly Glu Cys Ala Lys Gln Asp Leu Ala Thr Lys	620	625	630
Asn Ser Val Asp	Gly Ala Val Gly Ile Leu Val His Ile Trp Leu	635	640	645
Phe Pro Ile Ile	Val Leu Ile Ser Ile Leu Ser Pro Arg Arg	650	655	

<210> 11

<211> 626

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6717888CD1

<400> 11

Met Gly Pro Ala Trp	Val Gln Asp Pro Leu Thr Gly Ala Leu Trp	
1	5	10 15
Leu Pro Val Leu Trp	Ala Leu Leu Ser Gln Val Tyr Cys Phe His	
	20	25 30
Asp Pro Pro Gly Trp	Arg Phe Thr Ser Ser Glu Ile Val Ile Pro	
	35	40 45
Arg Lys Val Pro His	Arg Arg Gly Gly Val Glu Met Pro Asp Gln	
	50	55 60
Leu Ser Tyr Ser Met	His Phe Arg Gly Gln Arg His Val Ile His	
	65	70 75
Met Lys Leu Lys Lys	Asn Met Met Pro Arg His Leu Pro Val Phe	
	80	85 90
Thr Asn Asn Asp Gln	Gly Ala Met Gln Glu Asn Tyr Pro Phe Val	
	95	100 105
Pro Arg Asp Cys Tyr	Tyr Asp Cys Tyr Leu Glu Gly Val Pro Gly	
	110	115 120
Ser Val Ala Thr Leu	Asp Thr Cys Arg Gly Gly Leu Arg Gly Met	
	125	130 135
Leu Gln Val Asp Asp	Leu Thr Tyr Glu Ile Lys Pro Leu Glu Ala	
	140	145 150
Phe Ser Lys Phe Glu	Tyr Val Val Ser Leu Leu Val Ser Glu Glu	
	155	160 165
Arg Pro Gly Glu Val	Ser Arg Cys Lys Thr Glu Gly Glu Glu Ile	
	170	175 180
Asp Gln Glu Ser Glu	Lys Val Lys Leu Ala Glu Thr Pro Arg Glu	
	185	190 195
Gly His Val Tyr Leu	Trp Arg His His Arg Lys Asn Leu Lys Leu	
	200	205 210
His Tyr Thr Val Thr	Asn Gly Leu Phe Met Gln Asn Pro Asn Met	
	215	220 225
Ser His Ile Ile Glu	Asn Val Val Ile Ile Asn Ser Ile Ile His	
	230	235 240
Thr Ile Phe Lys Pro	Val Tyr Leu Asn Val Tyr Val Arg Val Leu	
	245	250 255
Cys Ile Trp Asn Asp	Met Asp Ile Val Met Tyr Asn Met Pro Ala	
	260	265 270
Asp Leu Val Val Gly	Glu Phe Gly Ser Trp Lys Tyr Tyr Glu Trp	
	275	280 285
Phe Ser Gln Ile Pro	His Asp Thr Ser Val Val Phe Thr Ser Asn	
	290	295 300
Arg Leu Gly Asn Thr	Pro Arg Cys Gly Asp Lys Ile Lys Asn Gln	
	305	310 315
Arg Glu Glu Cys Asp	Cys Gly Ser Leu Lys Asp Cys Ala Ser Asp	
	320	325 330
Arg Cys Cys Glu Thr	Ser Cys Thr Leu Ser Leu Gly Ser Val Cys	
	335	340 345
Asn Thr Gly Leu Cys	Cys His Lys Cys Lys Tyr Ala Ala Pro Gly	
	350	355 360
Val Val Cys Arg Asp	Leu Gly Gly Ile Cys Asp Leu Pro Glu Tyr	
	365	370 375
Cys Asp Gly Lys Lys	Glu Glu Cys Pro Asn Asp Ile Tyr Ile Gln	
	380	385 390
Asp Gly Thr Pro Cys	Ser Ala Val Ser Val Cys Ile Arg Gly Asn	
	395	400 405
Cys Ser Asp Arg Asp	Met Gln Cys Gln Ala Leu Phe Gly Tyr Gln	

410	415	420
Val Lys Asp Gly Ser Pro Ala Cys Tyr	Arg Lys Leu Asn Arg Ile	
425	430	435
Gly Asn Arg Phe Gly Asn Cys Gly Val	Ile Leu Arg Arg Gly Gly	
440	445	450
Ser Arg Pro Phe Pro Cys Glu Glu Asp	Asp Val Phe Cys Gly Met	
455	460	465
Leu His Cys Ser Arg Val Ser His Ile	Pro Gly Gly Gly Glu His	
470	475	480
Thr Thr Phe Cys Asn Ile Leu Val His	Asp Ile Lys Glu Glu Lys	
485	490	495
Cys Phe Gly Tyr Glu Ala His Gln Gly	Thr Asp Leu Pro Glu Met	
500	505	510
Gly Leu Val Val Asp Gly Ala Thr Cys	Gly Pro Gly Ser Tyr Cys	
515	520	525
Leu Lys Arg Asn Cys Thr Phe Tyr Gln	Asp Leu His Phe Glu Cys	
530	535	540
Asp Leu Lys Thr Cys Asn Tyr Lys Gly	Val Cys Asn Asn Lys Lys	
545	550	555
His Cys His Cys Leu His Glu Trp Gln	Pro Pro Thr Cys Glu Leu	
560	565	570
Arg Gly Lys Gly Gly Ser Ile Asp Ser	Gly Pro Leu Pro Asp Lys	
575	580	585
Gln Tyr Arg Ile Ala Gly Ser Ile Leu	Val Asn Thr Asn Arg Ala	
590	595	600
Leu Val Leu Ile Cys Ile Arg Tyr Ile	Leu Phe Val Val Ser Leu	
605	610	615
Leu Phe Gly Gly Phe Ser Gln Ala Ile	Gln Cys	
620	625	

<210> 12

<211> 557

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472044CD1

<400> 12

Met Leu Leu Ala Val Leu Leu Leu Leu Pro Leu Pro Ser Ser Trp	
1 5 10 15	
Phe Ala His Gly His Pro Leu Tyr Thr Arg Leu Pro Pro Ser Ala	
20 25 30	
Leu Gln Val Phe Thr Leu Leu Leu Gly Ala Glu Thr Val Leu Gly	
35 40 45	
Arg Asn Leu Asp Tyr Val Cys Glu Gly Pro Cys Gly Glu Arg Arg	
50 55 60	
Pro Ser Thr Ala Asn Val Thr Arg Ala His Gly Arg Ile Val Gly	
65 70 75	
Gly Ser Ala Ala Pro Pro Gly Ala Trp Pro Trp Leu Val Arg Leu	
80 85 90	
Gln Leu Gly Gly Gln Pro Leu Cys Gly Gly Val Leu Val Ala Ala	
95 100 105	
Ser Trp Val Leu Thr Ala Ala His Cys Phe Val Gly Cys Arg Ser	
110 115 120	
Thr Arg Ser Ala Pro Asn Glu Leu Leu Trp Thr Val Thr Leu Ala	
125 130 135	
Glu Gly Ser Arg Gly Glu Gln Ala Glu Glu Val Pro Val Asn Arg	
140 145 150	
Ile Leu Pro His Pro Lys Phe Asp Pro Arg Thr Phe His Asn Asp	
155 160 165	
Leu Ala Leu Val Gln Leu Trp Thr Pro Val Ser Pro Gly Gly Ser	

	170		175		180
Ala Arg Pro Val	Cys Leu Pro Gln Glu	Pro Gln Glu Pro Pro	Ala		
	185		190		195
Gly Thr Ala Cys	Ala Ile Ala Gly Trp	Gly Ala Leu Phe Glu	Asp		
	200		205		210
Gly Pro Glu Ala	Glu Ala Val Arg Glu	Ala Arg Val Pro Leu	Leu		
	215		220		225
Ser Thr Asp Thr	Cys Arg Arg Ala Leu	Gly Pro Gly Leu Arg	Pro		
	230		235		240
Ser Thr Met Leu	Cys Ala Gly Tyr Leu	Ala Gly Gly Val Asp	Ser		
	245		250		255
Cys Gln Gly Asp	Ser Gly Gly Pro Leu	Thr Cys Ser Glu Pro	Gly		
	260		265		270
Pro Arg Pro Arg	Glu Val Leu Phe Gly	Val Thr Ser Trp Gly	Asp		
	275		280		285
Gly Cys Gly Glu	Pro Gly Lys Pro Gly	Val Tyr Thr Arg Val	Ala		
	290		295		300
Val Phe Lys Asp	Trp Leu Gln Glu Gln	Met Ser Ala Ser Ser	Ser		
	305		310		315
Ser Arg Glu Pro	Ser Cys Arg Glu Leu	Leu Ala Trp Asp Pro	Pro		
	320		325		330
Gln Glu Leu Gln	Ala Asp Ala Ala Arg	Leu Cys Ala Phe Tyr	Ala		
	335		340		345
Arg Leu Cys Pro	Gly Ser Gln Gly Ala	Cys Ala Arg Leu Ala	His		
	350		355		360
Gln Gln Cys Leu	Gln Arg Arg Arg Arg	Cys Glu Leu Arg Ser	Leu		
	365		370		375
Ala His Thr Leu	Leu Gly Leu Leu Arg	Asn Ala Gln Glu Leu	Leu		
	380		385		390
Gly Pro Arg Pro	Gly Leu Arg Arg Leu	Ala Pro Ala Leu Ala	Leu		
	395		400		405
Pro Ala Pro Ala	Leu Arg Glu Ser Pro	Leu His Pro Ala Arg	Glu		
	410		415		420
Leu Arg Leu His	Ser Gly Cys Pro Gly	Leu Glu Pro Leu Arg	Gln		
	425		430		435
Lys Leu Ala Ala	Leu Gln Gly Ala His	Ala Trp Ile Leu Gln	Val		
	440		445		450
Pro Ser Glu His	Leu Ala Met Asn Phe	His Glu Val Leu Ala	Asp		
	455		460		465
Leu Gly Ser Lys	Thr Leu Thr Gly Leu	Phe Arg Ala Trp Val	Arg		
	470		475		480
Ala Gly Leu Gly	Gly Arg His Val Ala	Phe Ser Gly Leu Val	Gly		
	485		490		495
Leu Glu Pro Ala	Thr Leu Ala Arg Ser	Leu Pro Arg Leu Leu	Val		
	500		505		510
Gln Ala Leu Gln	Ala Phe Arg Val Ala	Ala Leu Ala Glu Gly	Glu		
	515		520		525
Pro Glu Gly Pro	Trp Met Asp Val Gly	Gln Gly Pro Gly Leu	Glu		
	530		535		540
Arg Lys Gly His	His Pro Leu Asn Pro	Gln Val Pro Pro Ala	Arg		
	545		550		555
Gln Pro					

<210> 13

<211> 494

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477384CD1

<400> 13

```

Met Gly Gly Pro Cys Arg Ala Pro Leu Gln Pro Gln Cys Ala Arg
 1          5          10          15
Arg Arg Glu Ala Trp Ala Arg Arg His Arg Arg Arg Gly Ala Gly
 20          25          30
Arg Arg Arg Arg Gly Gly Ala Pro Ala Ala Arg Ala Gly Arg Gly
 35          40          45
Arg Gly Arg Gly Arg Gly Ala Leu Arg Gly Pro Gly Arg Pro Trp
 50          55          60
Ala Pro Pro Pro Pro Ala Pro Arg Pro Ala Ala Gly Pro Ala Pro
 65          70          75
Pro Pro Thr Arg Ser Leu Ser Pro Pro Leu Arg Pro Ala Val Pro
 80          85          90
Pro Ser Arg Arg Arg Leu Phe Leu Gly Glu Ala Leu Phe Gln Arg
 95          100          105
Ala Gly Ser Met Ala Ala Val Glu Thr Arg Val Cys Glu Thr Asp
 110          115          120
Gly Cys Ser Ser Glu Ala Lys Leu Gln Cys Pro Thr Cys Ile Lys
 125          130          135
Leu Gly Ile Gln Gly Ser Tyr Phe Cys Ser Gln Glu Cys Phe Lys
 140          145          150
Gly Ser Trp Ala Thr His Lys Leu Leu His Lys Lys Ala Lys Asp
 155          160          165
Glu Lys Ala Lys Arg Glu Val Ser Ser Trp Thr Val Glu Gly Asp
 170          175          180
Ile Asn Thr Asp Pro Trp Ala Gly Tyr Arg Tyr Thr Gly Lys Leu
 185          190          195
Arg Pro His Tyr Pro Leu Met Pro Thr Arg Pro Val Pro Ser Tyr
 200          205          210
Ile Gln Arg Pro Asp Tyr Ala Asp His Pro Leu Gly Met Ser Glu
 215          220          225
Ser Glu Gln Ala Leu Lys Gly Thr Ser Gln Ile Lys Leu Leu Ser
 230          235          240
Ser Glu Asp Ile Glu Gly Met Arg Leu Val Cys Arg Leu Ala Arg
 245          250          255
Glu Val Leu Asp Val Ala Ala Gly Met Ile Lys Pro Gly Val Thr
 260          265          270
Thr Glu Glu Ile Asp His Ala Val His Leu Ala Cys Ile Ala Arg
 275          280          285
Asn Cys Tyr Pro Ser Pro Leu Asn Tyr Tyr Asn Phe Pro Lys Ser
 290          295          300
Cys Cys Thr Ser Val Asn Glu Val Ile Cys His Gly Ile Pro Asp
 305          310          315
Arg Arg Pro Leu Gln Glu Gly Asp Ile Val Asn Val Asp Ile Thr
 320          325          330
Leu Tyr Arg Asn Gly Tyr His Gly Asp Leu Asn Glu Thr Phe Phe
 335          340          345
Val Gly Glu Val Asp Asp Gly Ala Arg Lys Leu Val Gln Thr Thr
 350          355          360
Tyr Glu Cys Leu Met Gln Ala Ile Asp Ala Val Lys Pro Gly Val
 365          370          375
Arg Tyr Arg Glu Leu Gly Asn Ile Ile Gln Lys His Ala Gln Ala
 380          385          390
Asn Gly Phe Ser Val Val Arg Ser Tyr Cys Gly His Gly Ile His
 395          400          405
Lys Leu Phe His Thr Ala Pro Asn Val Pro His Tyr Ala Lys Asn
 410          415          420
Lys Ala Val Gly Val Met Lys Ser Gly His Val Phe Thr Ile Glu
 425          430          435
Pro Met Ile Cys Glu Gly Gly Trp Gln Asp Glu Thr Trp Pro Asp
 440          445          450
Gly Trp Thr Ala Val Thr Arg Asp Gly Lys Arg Ser Ala Gln Phe
 455          460          465

```

Glu	His	Thr	Leu	Leu	Val	Thr	Asp	Thr	Gly	Cys	Glu	Ile	Leu	Thr
				470					475					480
Arg	Arg	Leu	Asp	Ser	Ala	Arg	Pro	His	Phe	Met	Ser	Gln	Phe	
				485					490					

<210> 14

<211> 593

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7077175CD1

<400> 15

Met	Asn	Val	Leu	Lys	Leu	Asp	Thr	Leu	Val	Val	Ala	Gln	Leu	Trp
1				5					10					15
Arg	Tyr	Glu	Asn	Ala	Lys	Pro	Thr	Gly	Glu	Leu	Gly	Glu	Pro	Tyr
				20					25					30
Glu	Ala	Gly	Ile	Asn	Cys	Ser	Gly	Ser	Gly	Ala	Glu	Glu	Lys	Glu
				35					40					45
Asp	Arg	Arg	Met	Ala	Ile	Ile	Trp	Ala	Val	Pro	Ser	Thr	Ser	Val
				50					55					60
Ser	Trp	Glu	Gln	Thr	Ser	Arg	Lys	Thr	Gln	Ile	Arg	Lys	Lys	Arg
				65					70					75
Pro	Ala	Pro	Arg	Cys	Lys	Gln	Leu	Gly	Thr	Arg	Gln	Arg	Val	Leu
				80					85					90
Pro	Val	Val	Lys	Pro	Glu	Val	Leu	Gln	Lys	Ala	Thr	Val	Glu	Leu
				95					100					105
Leu	Asp	Gln	Ala	Leu	Cys	Ala	Ser	Leu	Tyr	Gly	His	Ser	Leu	Thr
				110					115					120
Asp	Arg	Met	Val	Cys	Ala	Gly	Tyr	Leu	Asp	Gly	Lys	Val	Asp	Ser
				125					130					135
Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Val	Cys	Glu	Glu	Pro	Ser
				140					145					150
Gly	Arg	Phe	Phe	Leu	Ala	Gly	Ile	Val	Ser	Trp	Gly	Ile	Gly	Cys
				155					160					165
Ala	Glu	Ala	Arg	Arg	Pro	Gly	Val	Tyr	Ala	Arg	Val	Thr	Arg	Leu
				170					175					180
Arg	Asp	Trp	Ile	Leu	Glu	Ala	Thr	Thr	Lys	Ala	Ser	Met	Pro	Leu
				185					190					195
Ala	Pro	Thr	Met	Ala	Pro	Ala	Pro	Ala	Ala	Pro	Ser	Thr	Ala	Trp
				200					205					210
Pro	Thr	Ser	Pro	Glu	Ser	Pro	Val	Val	Ser	Thr	Pro	Thr	Lys	Ser
				215					220					225
Met	Gln	Ala	Leu	Ser	Thr	Val	Pro	Leu	Asp	Trp	Val	Thr	Val	Pro
				230					235					240
Lys	Leu	Gln	Glu	Cys	Gly	Ala	Arg	Pro	Ala	Met	Glu	Lys	Pro	Thr
				245					250					255
Arg	Val	Val	Gly	Gly	Phe	Gly	Ala	Ala	Ser	Gly	Glu	Val	Pro	Trp
				260					265					270
Gln	Val	Ser	Leu	Lys	Glu	Gly	Ser	Arg	His	Phe	Cys	Gly	Ala	Thr
				275					280					285
Val	Ala	Gly	Asp	Arg	Trp	Leu	Leu	Ser	Ala	Ala	His	Cys	Phe	Asn
				290					295					300
His	Thr	Lys	Val	Glu	Gln	Val	Arg	Ala	His	Leu	Gly	Thr	Ala	Ser
				305					310					315
Leu	Leu	Gly	Leu	Gly	Gly	Ser	Pro	Val	Lys	Ile	Gly	Leu	Arg	Arg
				320					325					330
Val	Val	Leu	His	Pro	Leu	Tyr	Asn	Pro	Gly	Ile	Leu	Asp	Phe	Asp
				335					340					345
Leu	Ala	Val	Leu	Glu	Leu	Ala	Ser	Pro	Leu	Ala	Phe	Asn	Lys	Tyr
				350					355					360

Ile	Gln	Pro	Val	Cys	Leu	Pro	Leu	Ala	Ile	Gln	Lys	Phe	Pro	Val	365	370	375
Gly	Arg	Lys	Cys	Met	Ile	Ser	Gly	Trp	Gly	Asn	Thr	Gln	Glu	Gly	380	385	390
Asn	Ala	Thr	Lys	Pro	Glu	Leu	Leu	Gln	Lys	Ala	Ser	Val	Gly	Ile	395	400	405
Ile	Asp	Gln	Lys	Thr	Cys	Ser	Val	Leu	Tyr	Asn	Phe	Ser	Leu	Thr	410	415	420
Asp	Arg	Met	Ile	Cys	Ala	Gly	Phe	Leu	Glu	Gly	Lys	Val	Asp	Ser	425	430	435
Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Leu	Ala	Cys	Glu	Glu	Ala	Pro	440	445	450
Gly	Val	Phe	Tyr	Leu	Ala	Gly	Ile	Val	Ser	Trp	Gly	Ile	Gly	Cys	455	460	465
Ala	Gln	Val	Lys	Lys	Pro	Gly	Val	Tyr	Thr	Arg	Ile	Thr	Arg	Leu	470	475	480
Lys	Gly	Trp	Ile	Leu	Glu	Ile	Met	Ser	Ser	Gln	Pro	Leu	Pro	Met	485	490	495
Ser	Pro	Pro	Ser	Thr	Thr	Arg	Met	Leu	Ala	Thr	Thr	Ser	Pro	Arg	500	505	510
Thr	Thr	Ala	Gly	Leu	Thr	Val	Pro	Gly	Ala	Thr	Pro	Ser	Arg	Pro	515	520	525
Thr	Pro	Gly	Ala	Ala	Ser	Arg	Val	Thr	Gly	Gln	Pro	Ala	Asn	Ser	530	535	540
Thr	Leu	Ser	Ala	Val	Ser	Thr	Thr	Ala	Arg	Gly	Gln	Thr	Pro	Phe	545	550	555
Pro	Asp	Ala	Pro	Glu	Ala	Thr	Thr	His	Thr	Gln	Leu	Pro	Gly	Thr	560	565	570
Gly	Arg	Asp	Gly	Gly	Ile	Pro	Gly	Ser	Gly	Gly	Ser	His	Val	Asn	575	580	585
Gln	Pro	Gly	Leu	Pro	Asn	Lys	Thr								590		

<210> 15

<211> 319

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7480124CD1

<400> 16

Met	Gly	Pro	Leu	Gly	Pro	Ser	Ala	Leu	Gly	Leu	Leu	Leu	Leu	Leu	1	5	10	15
Leu	Val	Val	Ala	Pro	Pro	Arg	Val	Ala	Ala	Leu	Val	His	Arg	Gln	20	25	30	35
Pro	Glu	Asn	Gln	Gly	Ile	Ser	Leu	Thr	Gly	Ser	Val	Ala	Cys	Gly	40	45	50	55
Arg	Pro	Ser	Met	Glu	Gly	Lys	Ile	Leu	Gly	Gly	Val	Pro	Ala	Pro	60	65	70	75
Glu	Arg	Lys	Trp	Pro	Trp	Gln	Val	Ser	Val	His	Tyr	Ala	Gly	Leu	80	85	90	95
His	Val	Cys	Gly	Gly	Ser	Ile	Leu	Asn	Glu	Tyr	Trp	Val	Leu	Ser	100	105	110	115
Ala	Ala	His	Cys	Phe	His	Arg	Asp	Lys	Asn	Ile	Lys	Ile	Tyr	Asp	120	125	130	135
Met	Tyr	Val	Gly	Leu	Val	Asn	Leu	Arg	Val	Ala	Gly	Asn	His	Thr	140	145	150	

```

Thr Arg Ile Val Phe Ser Glu Ser Val Leu Pro Val Cys Leu Ala
155 160 165
Thr Pro Glu Val Asn Leu Thr Ser Ala Asn Cys Trp Ala Thr Gly
170 175 180
Trp Gly Leu Val Ser Lys Gln Gly Glu Thr Ser Asp Glu Leu Gln
185 190 195
Glu Val Gln Leu Pro Leu Ile Leu Glu Pro Trp Cys His Leu Leu
200 205 210
Tyr Gly His Met Ser Tyr Ile Met Pro Asp Met Leu Cys Ala Gly
215 220 225
Asp Ile Leu Asn Ala Lys Thr Val Cys Glu Gly Asp Ser Gly Gly
230 235 240
Pro Leu Val Cys Glu Phe Asn Arg Ser Trp Leu Gln Ile Gly Ile
245 250 255
Val Ser Trp Gly Arg Gly Cys Ser Asn Pro Leu Tyr Pro Gly Val
260 265 270
Tyr Ala Ser Val Ser Tyr Phe Ser Lys Trp Ile Cys Asp Asn Ile
275 280 285
Glu Ile Thr Pro Thr Pro Ala Gln Pro Ala Pro Ala Leu Ser Pro
290 295 300
Ala Leu Gly Pro Thr Leu Ser Val Leu Met Ala Met Leu Ala Gly
305 310 315
Trp Ser Val Leu

```

<210> 16
 <211> 2406
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6926819CB1

<400> 17

```

gttaagctga aaatgcacac agggctcctg taaatttctt ttcataaacc acccgcccag 60
ggcattaaat aggggtactta gttgatccga accctccagg gagacctccg acccttctct 120
tcgtagcccc cagctccctt cccccgggtc cactgaggca aggggactga gctgctccac 180
atgccaggag tcagcacgcc ggaaggcccc gccagcggc tggcgcagcc aatcgcacag 240
cgggcaagtg gtggggggcg gcctgcctgg gcggcaaggg ggccagcggg tctaggggct 300
ttacagggtc attagctgct ttcggggcgc cttaggcgac aggagactcc tggaccagc 360
acctgcccac tgtgcctgtc cacctgtggc tacagcagct gagaccccag tgggctaaaag 420
attggacagg ggcccaccag ggacccagca agtccttcag ctctgtgagt gagggtattt 480
ccggagtgcc aggcgcagc attcccagg ccgtgggggt ggacaggggg gctcgacccc 540
ggcaaatcag gcagaggcgc cccttgctcc ctgcaacatc gccacgtcc tggggccaca 600
gtgagcatga gcggaggggc ggagcaagag ccaggggacc tggcctgggt cccagccca 660
aagcctggga agctgcctac ccacccctgt gtgggcgcgg aactggggga ctctggcttc 720
cgggtggttc gccacctgat tcagtttatg ctctgtgagg ggagctggag tgttggcagg 780
actgcccac ctgcaggact gcaggactgc gggaacggcg gtagatgggt gctctccttc 840
ccagtttgtc ctgggaagac attcaataac tgtttcatta caaggggcat ttggaaaaca 900
tacttcacct tctgttgtgt attagccaag aacaagggtg gatgtgactt cccaattatt 960
ggggatccct ttgtcccttc ttgaaattag atgtcttcat tcttgagggt ttgcctggat 1020
gacctcagca caattggtac aaaacctggg ccaatgggtt cctagtcttc cggttgttgc 1080
cttaagcttc tcgcccatac ggtaccttcc tgtccttggt catagcctgt catcatcatt 1140
ccagaaaact gtttcaactc ctacagctgt ggacaggctg cttttcattt tggtagggct 1200
ctccaatacc tccacttgcc ctgtttttct ccagccacat ccttggcctc ttccacagtc 1260
cttaggtaaa tgcttggaag aataatttaa atatttttat tctaccatgg tggccctagt 1320
ttctcagggg gtagtaaaat ggctttttag gatcggtcta atcagatcct catttctttt 1380
cccttcctag atttttgaaa catgaatcct tcactcctcc tggctgcctt tttcctggga 1440
attgcctcag ctgctctaac acgtgaccac agttagagcg cacaatggac caagtggag 1500
gcaaagcaca agagattata tggcatgaac aggaaccact ggattagagt cctctgggag 1560
aaggacgtga agatgattga gcagcacaat cagggaataca gccaagggaa acacagcttc 1620
acaatggcca tgaacgcctt tggagacatg gtaagtgaag aattcaggca ggtgatgaat 1680

```

```

ggttttcaat accagaagca caggaagggg aaacagttcc aggaacgcct gcttcttgag 1740
atccccacat ctgtggactg gagagagaaa ggctacatga ctctgtgaa ggatcagcag 1800
ggtcagtgtg gctcttggtg ggcttttagt gcaactgggtg ctctggaagg gcagatgttc 1860
tgaaaaacag gcaacttat ctactgaat gagcagaatc tggtagactg ctctgggcct 1920
caaggcaatg aggggtgcaa tggtgacttc atggataatc ccttccggtg tgttcaggag 1980
aacggaggcc tggactctga ggcatcctat ccatatgaag gaaagggtta aacctgtagg 2040
tacaatccca agtattctgc tgctaatac actgggtttt tggacatccc ttacggggag 2100
aaggacctgg cgaaggcagt ggcaactgtg gggcccatct ctgttgctgt tggtgcaagc 2160
catgtcttct tccagttcta taaaaaaggg atttattttg agccacgctg tgacctgaa 2220
ggcctggatc atgctatgct ggtggttggc tacagctatg aaggagcaga ctacagataac 2280
aataaatatt ggctggtgaa gaacagctgg ggtaaaaact ggggcattga tggctacata 2340
aagatggcca aagaccggag gaacaactgt ggaattgcc aagcagccag ctacccact 2400
gtgtga 2406

```

<210> 17

<211> 1967

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473526CB1

<400> 18

```

cggacgcgtg ggcggacgcg tgggtgcccc ggcgcttaaa gaagcaaaat ctcttgtgca 60
ggagcagcag agactcctca ggaagactca ctggactgta cccaccacct gccatgtctc 120
tgtggccacc tttccgatgc agatggaagc tggcgccaag gtactctagg agggcgtctc 180
cacagcaacc ccaacaggac tttgaggccc tgctggcaga gtgcctgagg aatggctgcc 240
tctttgaaga caccagcttc ccggccaccc tgagctccat cggcagtgcc tccctgctgc 300
agaagctgcc accccgcctg cagtggaga gggccccgga gctgcacagc aatccccagt 360
tttattttgc caaggccaaa aggtggatc tgtgccagg gatagtagga gactgctgg 420
tcttggtgct tttgcaagct ctggccttgc accaggacat cctgagccgg gttgttcccc 480
tgaatcagag tttactgag aagtatgctg gcactctccg gttctggttc tggcactatg 540
ggaactgggt tctgtggtg atcgatgacc gtctgcctgt gaatgaggct ggccagctgg 600
tctttgtctc ctccacctat aagaacttgt tctggggagc acttctggaa aaggcctatg 660
ccaagctctc tggttcctat gaagacttgc agtcaggaca ggtgtctgaa gccctttag 720
acttcaactg aggggtgaca atgaccatca acctggcaga agcccatggc aacctctggg 780
acatcctcat cgaagccacc tacaacagaa ccctcattgg ctgccagacc cactcagggg 840
agaagattct ggagaatggg ctggtggaag gccatgccta tactctcaca ggaatcagga 900
aggtgacctg caaacataga cctgaatata tcgtcaagct acggaacccc tggggaaagg 960
tggaatggaa aggagactgg agtgacagtt caagtaaatg ggagctgtg agccccaagg 1020
agaagattct gcttctgagg aaagacaatg acggagaatt ctggatgac ctgcaggact 1080
ttaaaccaca tttcgtgctc ctggttatct gtaaaactgac ccaggcctg ttgagccagg 1140
aggcggccca gaagtggacg tacaccatgc gggaggggag atgggagaag cggagcacag 1200
ctggtggcca gaggcagttg ctgcaggaca cattttggaa gaaccgcag tctctgtgt 1260
ctgtctggag gcccgaggag ggcaggagat ccctgaggcc ctgcagcgtg ctggtgtccc 1320
tgetccagaa gcccaggcac aggtgccgca agcgggaagc tctcctcgcc attggcttct 1380
acctctatag gatgaacaag taccatgat accagaggag actgccccct gagttcttcc 1440
agagaaacac tcctctgagc cagcctgata ggtttctcaa ggagaaagaa gtgagtcagg 1500
agctgtgtct ggaaccaggg acgtacctca tcgtgcctgc atattggagg cccaccagaa 1560
gtcagagttc gtcctcaggg tcttctccag gaagcacatc ttttatgaa ttggcagcaa 1620
ttctggtgtc gtcttctcaa aggagataga agacaaaat gaaaggcagg atgaattctt 1680
caccaaattc ttttgaaaag catccagaga ttaatgcagt tcaacttcag aacctcctga 1740
accagatgac ctgggtcaagt ctggggagca gacagccctt tctttagcct ggaagcctgc 1800
aggggatcct ggccttactg accttaatgc atcaggtact atgagcatcc caggaatcag 1860
gcacctgttg gaaggagtga agtctctcag aaggtctcca caagcaacac cgtgggtcag 1920
gaactgaact ggagcaatgg acgtgcagaa ggagcagaac acgccag 1967

```

<210> 18

<211> 3446

<212> DNA

<213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478443CB1

<400> 19

tgcttagagg	ccgaggagct	cacagctatg	ggctggaggc	cccggagagc	tcgggggacc	60
ccgttgctgc	tgctgctact	actgctgctg	ctctggccag	tgccaggcgc	cgggggtgctt	120
caaggacata	tccctgggca	gccagtcacc	ccgcactggg	tccctggatgg	acaaccctgg	180
cgcaccgtca	gcctggagga	gccggtctcg	aagccagaca	tggggctggg	ggccctggag	240
gctgaaggcc	aggagctcct	gcttgagctg	gagaagaacc	acaggctgct	ggccccagga	300
tacatagaaa	cccactacgg	cccagatggg	cagccagtg	tgctggcccc	caaccacacg	360
gatcattgcc	actaccaagg	gcgagtaagg	ggcttccccg	actcctgggt	agtctctctg	420
acctgctctg	ggatgagtg	cctgatcacc	ctcagcagga	atgccagcta	ttatctgcgt	480
ccctggccac	cccggggctc	caaggacttc	tcaaccacg	agatctttcg	gatggagcag	540
ctgctcacct	ggaaaggaa	ctgtggccac	agggatcctg	ggaacaaagc	gggcatgacc	600
agccttctctg	gtgggtccca	gagcaggggc	aggcgagaag	cgcgcaggac	ccggaagtac	660
ctggaactgt	acattgtggc	agaccacacc	ctgttcttga	ctcggcaccg	aaacttgaac	720
cacaccaaac	agcgtctcct	ggaagtgcgc	aactacgtgg	accagcttct	caggactctg	780
gacattcagg	tggtcgctgac	cggcctggag	gtgtggaccg	agcgggaccg	cagccgcgtc	840
acgcaggacg	ccaacgccac	gctctggggc	tctctgcagt	ggcgccgggg	gctgtggggg	900
cagcgggccc	acgactccgc	gcagctgctc	acgggcccgc	ccttccaggg	cggcacagt	960
ggcctggcgc	ccgtcgaggg	catgtgccgc	gccgagagct	cgggaggcgt	gagcacggac	1020
cactcggagc	tccccatcgg	cgcgcagacc	accatggccc	atgagatcgg	ccacagcctc	1080
ggcctcagcc	acgaccccca	cggctgctgc	gtggaggctg	cggccgagtc	cggaggctgc	1140
gtcatggctg	cggccaccgg	gcacccgttt	ccgcgcgtgt	tcagcgcctg	cagccgcctc	1200
cagctgcgcg	ccttcttcgg	caaggggggg	ggcgcttgcc	tctccaatgc	cccggacccc	1260
ggactccccc	tgccgcccgc	gctctgcggg	aacggcttcg	tggaagcggg	caggagtggt	1320
gactcggggc	ctggccagga	gtgccgcgac	ctctgctgct	ttgtccacaa	ctgctcgctg	1380
cgcgccgggg	cccagtgccg	ccacggggac	tgctgcgtgc	gctgcctgct	gaagccggct	1440
ggagcgctgt	gccgccaggc	catgggtgac	tgtgacctcc	ctgagttttg	cacgggcacc	1500
tctctccact	gtccccaga	cgtttaccta	ctggacggct	caccctgtgc	caggggcagt	1560
ggctactgct	gggatggcgc	atgtcccacg	ctggagcagc	agtgccagca	gctctggggg	1620
cctggctccc	accagctcc	cagggcctgt	ttccagggtg	tgaactctgc	gggagatgct	1680
catggaaact	gcggccagga	cagcgagggc	cacttctctg	cctgtgcagg	gagggatgcc	1740
ctgtgtggga	agctgcagt	ccagggtgga	aagcccagcc	tgctcgcacc	gcacatgggt	1800
ccagtggact	ctaccgttca	cctagatggc	caggaagtga	cttgtcgggg	agccttgga	1860
ctccccagt	cccagctgga	cctgcttggc	ctgggcctgg	tagagccagg	caccagtggt	1920
ggactcagaa	tggtgtgcca	gagcaggcgc	tgagggaaga	atgccttcca	ggagcttcag	1980
cgtgcctga	ctgcctgcca	cagccacggg	gtttgcaata	gcaaccataa	ctgccactgt	2040
gctccaggct	gggctccacc	cttctgtgac	aagccaggct	ttgggtggcag	catggacagt	2100
ggcctgtgct	aggctgaaaa	ccatgacacc	ttcctgctgg	ccatgtcctc	cagcgtcctg	2160
ctgcctctgc	tcccaggggc	cggcctggcc	tggtgttgc	accgactccc	aggagccccc	2220
ctgcagcgat	cagctggggg	ctgcagaagg	gacccctgcg	gcagtggccc	caaagatggc	2280
ccacacaggg	accacccctc	gggcccgtt	caccccatgg	agttggggcc	cacagccact	2340
ggacagccct	ggccctgga	ccctgagaac	tctcatgagc	ccagcagcca	ccctgagaag	2400
cctctgccag	cagtctcgcc	tgacccccaa	gatcaagtcc	agatgccaa	atcctgcctc	2460
tggtgagagg	tagctcctaa	aatgaacaga	tttaaagaca	ggtggccact	gacagccact	2520
ccaggaaact	gaactgcagg	ggcagagcca	gtgaatcacc	ggacctccag	cacctgcagg	2580
cagcttgaaa	gtttcttccc	cagtgaggag	ttcgacccac	ccactccagg	aacccagagc	2640
cacattagaa	gttctctgag	gctggagaac	actgctgggc	acactctcca	gctcaataaa	2700
ccatcagtc	cagaagcaaa	ggtcacacag	cccctgacct	ccctcaccag	tggaggctgg	2760
gtagtgctgg	ccatcccaaa	agggtctgt	cctgggagtc	tggtgtgtct	cctacatgca	2820
atttccacgg	accagctct	gtggagggca	tgaactgctg	ccagaagcta	gtggtcctgg	2880
ggcctatgg	ttcgactgag	tccacactcc	cctggagcct	ggctggcctc	tgcaaacaaa	2940
cataattttg	gggaccttcc	ttcctgtttc	ttcccacctc	gtcttctccc	ctagggtggt	3000
cctgagcccc	cacccccaat	cccagtgcct	cacctgaggt	tctggagctc	agaatctgac	3060
agcctctccc	ccattctgtg	tgtgtcgggg	ggacagaggg	aaccatttaa	gaaaagatac	3120
caaagtagaa	gtcaaaagaa	agacatgttg	gctataggcg	tggtggctca	tgcctataat	3180
cccagcactt	tggaagccg	gggtaggagg	atcacagag	gccagcagg	ccacaccagc	3240
ctgggcaaca	gcgaagaca	ccgcatctac	agaaaaattt	taaaattagc	tgggcgtggg	3300
ggtgtgtacc	tgtaggccta	gctgctcagg	aggctgaagc	aggaggatca	cttgagcctg	3360
agttcaacac	tgagtgagc	tatggtggca	ccactgcact	ccagcctggg	tgacagagca	3420
agacctgtgc	tctaaaataa	atttta				3446

<210> 19
 <211> 4888
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3533147CB1

<400> 20

```

atgacaggaa caggagggcag gaagcccact ggggacaaac aggaagtcca cccctgggaa 60
aaacaggaag tgaggggaaca gacagaaagt ccacaggagc tgacaagaag tccacagggg 120
acagacagga atgatacagt gaccatctat actgacaccc aaagccgaaa ggctggcgct 180
tctcgtaaaa tcagaaacat gctcaacatt taccttgttt ggtagttaaa gataaaccag 240
ataataatca atgtctttta tcaaaatcca gaaccaacta tctggaattc tgcattttatt 300
gtggacataa cagcaatagt tccaacagca ttatttccat ttaatgtggc caagccaaaa 360
atgctcgtgg agaattttaca ggaaggtgac ttcaggagc ttcgtggtaa cagccaccac 420
tgcttgacca aaaaggggtct aggaaatgct cctccaggcc tgcagttcac actgtacaaa 480
tgtctggact catccaggac agcccagccc catgcagggc ttcactacgt ggacattaat 540
tcaggcatga tacgaacaga agaggcagat tacttcctaa ggccacttcc ttcacacctc 600
tcatggaaac tcgceagagc tgcccaggc agctcgccat cccacgtact gtacaagaga 660
tccacagagc cccatgctcc tggggccagt gaggtcctgg tgacctcaag gacatgggag 720
ctggcacatc aacccctgca cagcagcgac cttcgctgg gactgccaca aaagcagcat 780
ttctgtggaa gacgcaagaa atacatgccc cagcctccca aggaagacct cttcatcttg 840
ccagatagt ataagtcttg cttacggcat aagcgctctc ttctgaggtc ccatagaaat 900
gaagaactga acgtggagac cttggtgggtg gtcgacaaaa agatgatgca aaacctggc 960
catgaaaata tcaccacctc cgtgctcacg atactcaaca tggatctgc tttattcaaa 1020
gatggaacaa taggagggaaa catcaacatt gcaattgtag gtctgattct tctagaagat 1080
gaacagccag gactgggtgat aagtccaccac gcagaccaca ccttaagtag cttctgccag 1140
tggcagctcg gatgtagggg gaaagatggg actcgctatg accacgccat cttactgact 1200
ggctctggata tatgttctcg gaagaatgag ccctgtgaca ctttgggatt tgcaccata 1260
agtggaatgt gtagtaataa tcgcagctgc acgattaatg aagatacagg tcttggaactg 1320
gccttcacca ttgcccataa gtctggacac aactttggca tgattcatga tggagaaggg 1380
aacatgtgta aaaagtcgga gggcaacatc atgtccccta cattggcagg acgcaatgga 1440
gtcttctcct ggtcaccctg cagccgccag tatctacaca aatttctaag caccgctcaa 1500
gctatctgcc ttgctgatca gccaaagcct gtgaaggaa acaagtatcc tgagaaattg 1560
ccaggagaat tatatgatgc aaacacacag tgcaagtggc agttcgaga gaaagccaag 1620
ctctgcatgc tggactttta aaaggacatc tgtaaggccc tgtggtgcca tctgtattga 1680
aggaaatgtg agactaaatt tatgccagca gcagaaggca caatttgtgg gcatgacatg 1740
tgggtgccggg gaggacagtg tgtgaaatat ggtgatgaag gccccaagcc caccatggc 1800
cactggctcg actggtcttc ttggtcccca tgctccagga cctgcggagg gggagtatct 1860
cataggagtc gcctctgcac caacccaag ccatcgcatg gaggaagtt ctgtgagggc 1920
tccactcgca ctctgaagct ctgcaacagt cagaaatgtc cccgggacag tgttgacttc 1980
cgtgctgctc agtgtgccga gcacaacagc agacgattca gagggcgcca ctacaagtgg 2040
aagccttaca ctcaagtaga agatcaggac ttatgcaaac tctactgtat cgcagaagga 2100
tttgatttct tcttttcttt gtcaataaaa gtcaaagatg ggactccatg ctcgagggat 2160
agccgtaatg tttgtataga tgggatattg gagagagttg gatgtgacaa tgtccttgga 2220
tctgatgctg ttgaagacgt ctgtgggggtg tgtaaccgga ataactcagc ctgacagatt 2280
cacaggggtc tctacaccaa gcaccaccac accaaccagt attatcacat ggtcaccatt 2340
ccttctggag cccggagtat ccgcatctat gaaatgaacg tctctacctc ctacatttct 2400
gtgcgcaatg cctcagaag gtactacctg aatgggcact ggaccgtgga ctggccggcg 2460
cggtacaaat tttcgggcac tactttcgac tacagacggc cctataatga gcccgagaac 2520
ttaatcgcta ctggaccaac caacgagaca ctgattgtgg agctgctgtt tcaggaagg 2580
aaccgggtg ttgctggga atactccatg cctcgcttgg ggaccgagaa gcagccctc 2640
gccagccca gctacacttg ggccatcgtg cgtctgagt gctccgtgtc ctgaggagg 2700
ggacagatga ccgtgagaga gggctgctac agagacctga agtttcaagt aaatatgtcc 2760
ttctgcaatc ccaagacacg acctgtcacg gggctgggtg cttgcaaagt atctgcctgt 2820
cctccagct ggtccgtggg gaactggagt gcctgcagtc ggacgtgtgg cgggggtgcc 2880
cagagccgcc ccgtgcagtg cacacggcgg gtgcacatg actcggagcc agtcccggc 2940
ggcctgtgcc ctcagctggt ccctccagca ggcaggcctg caactctcag agctgccac 3000
ctgcatggag cgccgggccc tgggcagagt gctcacacac ctgtgggaag ggtggaggaa 3060
cgggcagtg cctgtaagag caccaacccc tcggccagag cgagctgct gcccgacgct 3120
gtctgcacct ccgagcccaa gccaggatg catgaagcct gtctgcttca gcgtgccac 3180

```

```

aagcccaaga agctgcagtg gctgggtgtcc gcttgggtccc agtgctctgt gacatgtgaa 3240
agaggaacac agaaaagatt cttaaaatgt gctgaaaagt atgtttctgg aaagtatcga 3300
gagctggcct caaagaagtg ctacacattt ccgaagccca gcctggagct ggaacgtgcc 3360
tgcccccgcg ttccatgccc caggcaccgc ccatttctgt ctgcgggacc ctcgaggggc 3420
agctggtttg cctcaccctg gtctcagtg acggccagct gtgggggagg cgttcagacg 3480
aggctcgtgc agtgccctggc tgggggcggc ccggcctcag gctgcctcct gcaccagaag 3540
ccttcggcct ccctggcctg caacactcac ttctgcccc a ttgcagagaa gaaagatgcc 3600
ttctgcaaag actacttcca ctggtgctac ctggtacccc agcacgggat gtgcagccac 3660
aagttctacg gcaagcagtg ctgcaagact tgctctaagt ccaacttgtg agttgggacc 3720
gctctccgta gcagagaaag tgctgctggt gcacagaaat ttcccacaaa tgagctgtgc 3780
aatctacgtc ggaatacatc caaggaagag caaagccaaa agaagaaaac cgtgttaggc 3840
tctttgacca ggagtgtatg tatgtggttc actgtgagcc tgggtgcaga cctgtgtccc 3900
catgcacaca gtgtctcctg tcaggctgaa atgtggcacc ctggcagaca gagctgtggc 3960
tcgtgaggca gaaggcaggc accacaacgg gagaggcagc actcaccctt gcctgttgca 4020
gctaaatcaa gtcaaaaaga caggcgaggc tgaacttgt aaatgtctgg tgccttagaa 4080
aaagaaggaa ccgcatgaa ataaggaaaa catacaaaat atgtaccccc tagttcacca 4140
gcctcccctc ccactaggag ggccccctga gccatcagga gtgaccaact tcctgggtgg 4200
aggctcaggg agctccagga ggctgccag gctcctcctc ctccctccca gcggccgagc 4260
atctcttacc aggaacctgg agccaccgcc ggagccagcg tcatctctag ggctcactgc 4320
caggggactg cattctggtt tgggactttg cctatggaaa tgggaaaaat gaaattctctg 4380
ctaaggtgct tctatctctt tcagattcat gcattgaagg agagattttt tatactttat 4440
gttttatctt tctcagttat ttgcaagtga gtgtcctttt aaaaacacac ttcttcatgc 4500
ttttctttgt aaatgacaga tcgaagtata gggtacatca aaacctacc atcctgagaa 4560
gagttatggg tctattatag cagacgtcag ccacacagcc tatgtgacaa taaccttaga 4620
gtcctgtgtt ttgtttttgt gtgtgtgtgag attttaatct tttttttttt cgggtgagtct 4680
ggccatttct ataatgccag gtgggaagcc aggtgcggg tgttaggggt ggaatctgac 4740
cgcgctctct ggcacctcc ctgccatcct cagtgcggct gctgttctcc tgtccgggtc 4800
tgtggctcca ttccaaaggg gcacctggat atttatat tt gctgaagttt tataataaag 4860
tttatatggt acagtgaaaa aaaaaaaaaa 4888

```

<210> 20

<211> 1074

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7483438CB1

<400> 21

```

gccgtgcaca acaccaagcg catggactct gaccccagtg cagtgtctgt aggggaagagg 60
agccatgggg cttegggcag gcccatcct gcttctgtct ctgtggctgc tgccaggggc 120
ccattgggat gtgtgcctt cagaatgcgg ccactccaag gaggcgggga ggattgtggg 180
aggccaagac acccaggaag gacgctggcc gtggcaggtt ggctgtggt tgacctcagt 240
ggggcatgta tgtgggggct ccctcatcca ccacgctgg gtgtcacag ccgcccactg 300
cttcctgagg tctgaggatc ccgggctcta ccatgttaaa gtcggagggc tgacaccctc 360
actttcagag cccactcgg ccttgggtggc tgtgaggagg ctccgtgtcc actcctcata 420
ccatgggacc accaccagcg gggacattgc cctgatggag ctggactccc ccttgcaggc 480
ctcccagttc agccccatct gctcccagg accccagacc cccctcgcca ttgggaccgt 540
gtgctgggta aacgggctgg gggagggtggc tgtgcccctc ctggactcga acatgtgtga 600
gctgatgtac cacctaggag agcccagcct ggctggccag cgcctcatcc aggacgacat 660
gctctgtgct ggctctgtcc agggcaagaa agactcctgc cagggtgact ccgggggggc 720
gctggtctgc cccatcaatg atacgtggat ccaggccggc attgtgagct ggggattcgg 780
ctgtgccccg cttttccggc ctggtgtcta caccagggtg ctaagctaca cagactggat 840
tcagagaacc ctggctgaat ctactcagg catgtctggg gcccgccag gtgccccagg 900
atccactca ggcacctcca gatccacccc agtgctgctg cttgagctgt tgaccgtatg 960
cttgcttggg tccctgtgaa ccatgagcca tggagtccgg gatcccttt ctggtaggat 1020
tgatggaatc taataataaa aactgtaggt tttttatgtg taataaaaaa aaaa 1074

```

<210> 21

<211> 3573

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7246467CB1

<400> 22

```
caagaattcg gcacaggggt tctgtatccc tacttccatt accctcgggt cctcccactc 60
ctcgggggct ccgtgctttc cgcgggtctg tccgggggct ccggaccctc ggcgacgtg 120
agttgatggc ttccggagaa ctggcatagc tgcagaatat gagtagtgtc ccaagaagag 180
tgctttgctt ttgcgacaag gatcagaata aagttatttg ccatacagta accagagacc 240
tccaaactagg ggccccc aaa ctgtatcctg cctgtagaac ctgccaggta aaggtatatt 300
tttgcttttt aatttagcca gaagcaattt ttaaagaaaa tatgtctcct ctgaagatac 360
atggctctat cagaattcga agtatgcaga ctgggattac aaagtggaaa gaaggatcct 420
ttgaaattgt agaaaaagag aataaagtca gcctagtagt tcaactacaat actggaggaa 480
ttccaaggat atttcagcta agtcataaca ttaaaaatgt ggtgcttcga cccagtggag 540
cgaaacaaag ccgcctaag ttaactctgc aagataacag cttcttgtct attgacaaag 600
taccaagtaa ggtgacagag gaaatgaggt tgtttctaga tgcagtccat caaaacagac 660
ttcttgacgc catgaaaccg tctcaggggt ctggtagttt tggagccatt ctgggcagca 720
ggacctcaca gaaggaaacc agcaggcagc tttcttactc agacaatcag gcttctgcaa 780
aaagaggaag tttggaaact aaagatgata ttccatttcg aaaagttcct ggtaatccgg 840
gtagaggatc gattaagact gtagcaggaa gtggaatagc tcggacgatt ccttctttga 900
catctacttc aacacctctt agatcagggt tgctagaaaa tcgtactgaa aagaggaaaa 960
gaatgatatc aactggctca gaattgaatg aagattaccc taaggaaaaat gattcatcat 1020
cgaacaacaa ggccatgaca gatccctcca gaaagtattt aaccagcagt agagaaaagc 1080
agctgagttt gaaacagtca gaagagaata ggacatcagg tgggctttta cctttacagt 1140
catcatcctt ttatggtagc agagctggat ccaaggaaca ctcttctggt ggactaact 1200
tagacaggac taatgtttca agccagactc cctctgccaa aagaagtttg ggatttcttc 1260
ctcagccagt tctctttct gttaaaaaac tgagggtgtaa ccaggattac actggctgga 1320
ataaaccaag agtgccccct tctctcacc aacagcagca actgcagggc ttctccaatt 1380
tgggaaatac ctgctatatg aatgctattc tacaatctct attttcactc cagtcatctg 1440
caaatgactt gcttaaacaa ggtatcccat ggaagaaaat tccactcaat gcacttatca 1500
gacgctttgc acacttgctt gttaaaaaag atatctgtaa ttcagagacc aaaaaggatt 1560
tactcaagaa gggttaaaaat gccatttcag ctacagcaga gagattctct ggttatatgc 1620
agaatgatgc tcatgaattt ttaagtcagt gtttgacca gctgaaagaa gatatggaaa 1680
aattaaataa aacttggaag actgaacctg tttctggaga agaaaattca ccagatattt 1740
cagctaccag agcatacact tgccctgtta ttactaattt ggagtttgag gttcagcact 1800
ccatcatttg taaagcatgt ggagagatta tccccaaaag agaacagttt aatgacctct 1860
ctattgacct tctcgtagg aaaaaaccac tccctcctcg ttcaattcaa gattctcttg 1920
atcttttctt tagggccgaa gaactggagt attcttgtga gaagtgtggt gggaagtgtg 1980
ctcttgtcag gcacaaattt aacaggcttc ctagggtcct cattctccat ttgaaacgat 2040
atagcttcaa tgtggctctc tcgcttaaca ataagattgg gcagcaagtc atcattccaa 2100
gatactgac cctgtcatct cattgcactg aaaatacaaa accacctttt acccttgggt 2160
ggagtgcaca tatggcaatg tctagaccat tgaaagcctc tcaaatgggtg aattcctgca 2220
tcaccagccc ttctacacct tcaaaagaaat tcaccttcaa atccaagagc tccttggctt 2280
tatgccttga ttcagacagt gaggatgagc taaaacgttc tgtggccctc agccagagac 2340
tttgtgaaat gttaggcaac gaacagcagc aggaagacct ggaaaaagat tcaaaattat 2400
gcccaataga gcctgacaag tctgaattgg aaaactcagg atttgacaga atgagcgaag 2460
aagagcttct agcagctgtc ttggagataa gtaagagaga tgcttcacca tctctgagtc 2520
atgaagatga tgataagcca actagcagcc agataccgg atttgacaga gatgatattc 2580
aagaaatgcc agaaaatcca gacactatgg aaactgagaa gcccaaaaca atcacagagc 2640
tggaacctgc cagttttact gagataacta aagactgtga tgagaataaa gaaaacaaaa 2700
ctccagaagg atctcaggga gaagttgatt ggctccagca gtatgatatg gagcgtgaaa 2760
gggaagagca agagcttcag caggcactgg ctacagacct tcaagagcaa gaggcttggg 2820
aacagaaaga agatgatgac ctcaaaagag ctaccgagtt aagtcttcaa gaggtttaaca 2880
actccttgtt ggttctgatg ggttctgatg aggactctgg aaatgaggat gtttttgata 2940
tggaagacac agaagctgaa gctgaggaac tgaaaagaaa tgctgagaca ggaaatctgc 3000
ctcattcgta ccggctcatc agtggtgtca gtcacattgg tagcacttct tcttcaggtc 3060
attacattag tgatgtatat gacattaaga agcaagcgtg gtttacttac aatgacctgg 3120
aggtatcaaa aatccaagag gctgccgtgc agagtgatcg agatcggagt ggctacatct 3180
tcttttatat gcacaaggag atctttgatg agctgctgga aacagaaaag aactctcagt 3240
cacttagcac ggaagtggg aagactaccc gtcaggcctc gtgaggaaca aactcctggg 3300
ttggcagcat gcaactgcata tttgttactg ctgccacct cacctttcct ctgctgaagg 3360
agaatttggg attctacttg atgcgggagc aacaaacagc tcagggccaa accaaaagac 3420
aaaaattgga gtaacgtaga atgctccatg ctattttatg gaaacttttg tctcacatcc 3480
```

gtagctgatt atcctctttt tctcctatga gtggcacttc ttttgtctta ggaatacctg 3540
 ttgtacatct gtctccgtgt tgtgtttttt ccc 3573

<210> 22

<211> 4659

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7997881CB1

<400> 23

```

ggcggcgggc gcggcgctga cccggaggcg gcggcggcgg tgcccggatg gaggcacgtc 60
attgtacccc cgccgggggg ctgggctgtg tgcggcgcg cgccggcgcg cccgagggga 120
tggagcgagc gccgagccgg gtcagagttg aacaatgacc atagttgaca aagcttctga 180
atcttcagac ccacagcct atcagaatca gcctggcagc tccgaggcag tctcacctgg 240
agacatggat gcaggttctg ccagctgggg tgctgtgtct tcattgaatg atgtgtcaaa 300
tcacacactt tctttaggac cagtacctgg tgctgtagt tattcgagtt catctgtacc 360
tgataaatca aaaccatcac cacaaaagga tcaagcccta ggtgatggca tcgctcctcc 420
acagaaagtt cttttcccat ctgagaagat ttgtcttaag tggcaacaaa ctcatagagt 480
tggagctggg ctccagaatt tgggcaatac ctgttttgcc aatgcagcac tgcagtgttt 540
aacctacaca ccacctcttg ccaattacat gctatcacat gaacactcca aaacatgtca 600
tgcagaaggc ttttgtatga tgtgtacaat gcaagcacat attaccagg cactcagtaa 660
tcctggggac gttattaaac caatgtttgt catcaatgag atgcggcgta tagctaggca 720
cttccgtttt ggaaaccaag aagatgccca tgaattcctt caatacactg ttgatgctat 780
gcagaaagca tgcttgaatg gcagcaataa attagacaga cacaccagg ccaccactct 840
tgtttgcag atatttggag gatacctaag atctagagtc aaatgtttaa attgcaaggg 900
cgtttcagat acttttgatc catatcttga tataacattg gagataaagg ctgctcagag 960
tgtcaacaag gcattggagc agtttgtgaa gccggaacag cttgatggag aaaactcgta 1020
caagtgcagc aagtgtaaaa agatggttcc agcttcaaag aggttcacta tccatagatc 1080
ctctaattgt cttacacttt ctctgaaacg ttttgcaaat tttaccggtg gaaaaattgc 1140
taaggatgtg aaataccctg agtatcttga tatteggcca tatatgtctc aacccaacgg 1200
agagccaatt gtctacgtct tgtatgcagt gctggtccac actggtttta attgccatgc 1260
tggccattac ttctgtaca taaaagctag caatggcctc tggatatcaa tgaatgactc 1320
cattgtatct accagtgata ttagatcggg actcagccaa caagcctatg tgctctttta 1380
tatcaggtcc catgatgtga aaaaatggagg tgaacttact catcccacc atagccccgg 1440
ccagtcctct ccccgccccg tcatacagtc gcgggttgtc accaacaac aggtgcgccc 1500
aggctttatc ggaccacagc ttccctctca catgataaag aatccacctc acttaaatgg 1560
gactggacca ttgaaagaca cgccaagcag ttccatgtcg agtcctaacg ggaattccag 1620
tgtcaacagg gctagtcctg ttaatgctt agcttctgtc caaaactggt cagttaatat 1680
gtcctcagtg atcccagaac atcctaagaa acaaaaaatt acaatcagta ttcacaacaa 1740
gttgctgtt cgccagtgtc agtctcaacc taaccttcat agtaattctt tggagaaccc 1800
taccagccc gtccctctt ctaccattac caattctgca gtacagtcta cctcgaacgc 1860
atctacgatg tcagtttcta gtaaagtaac aaaaccgatc ccccgcatg aatcctgtct 1920
ccagcccgtg atgaatggca aatccaagct gaactccagc gtgctggtgc cctatggcgc 1980
cgagtctctc gaggactctg acgaggagtc aaaggggctg ggcaaggaga atgggattgg 2040
tacgatttgt agtcccact ctcccggcca agatgccgaa gatgaggagg cactccgca 2100
cgagcttcaa gaacccatga ccctaaacgg tgctaatagt gcagacagcg acagtgaacc 2160
gaaagaaaac ggcttagcgc ctgatggtgc cagctgcaa ggccagcctg ccctgcactc 2220
agaaaatccc tttgctaagg caaacggtct tcctggaaag ttgatgcctg ctcttttget 2280
gtctctccca gaagacaaaa tcttagagac cttcaggctt agcaacaaac tgaaaggctc 2340
gacggatgaa atgagtgcac ctggagcaga gaggggccct cccgaggacc gcgacgccga 2400
gcctcagcct ggagccccg ccgccgaatc cctggaggag ccagatgcgg ccgcccggcct 2460
cagcagcacc aagaaggctc cgccgccccg cgatccccgc acccccgcta ccaaagaagg 2520
cgctggggag gccatggccg tcgcccccg ggagcctccg cccagcgccg gcgaggacat 2580
cgtgggggac acagcaccct ctgacctgtg tgatcccggt agcttaacag gcgatgcgag 2640
cccgttgtcc caggacgcaa aggggatgat cgcgaggggc ccgcgggact cggcgttggc 2700
ggaagccccg gaagggttga gtccggctcc gcctgcgcgg tcggaggagc cctgcgagca 2760
gccactcctt gttcaccaca gcgggggacc cgcccgggac gctcaggacc catcccagag 2820
cttggggcga cccgaggccg cagagcggcc gccagctcct gtgctggaca tggccccggc 2880
cggtcacccc gaaggggacg ctgagcctag ccccggcgag agggctcagag acgcccgggc 2940
gccgaaagcc ccaggccctt ccccgcgcaa ggagaaaatc ggcagcctca gaaaggtgga 3000

```

```

ccgaggccac taccgcagcc ggagagagcg ctcgtccagc ggggagcccc ccagagagag 3060
caggagcaag actgagggcc accgtcaacc gcggcgcgcg acctgcccc gggagcgga 3120
cgccaggac cgccacgccc cggagcacca cccggccac ggcgacaggc tcagccctgg 3180
cgagcgccgc tctctgggca ggtgcagtc ccaccactcc cgacaccgga gcgggggtgga 3240
gctggactgg gtcagacacc actacaccga gggcgagcgt ggctggggcc gggagaagt 3300
ctaccccgac aggcgcgcgt gggacaggtg ccggtactac catgacaggt acgcctgtga 3360
cgctgcccgg gactggaagc ccttccacgg cggccgcgag cagcagcggg ccgggtgca 3420
cgagcgccg cacaaggacc acaaccgggg ccgtaggggc tgcgagccgg cccgggagag 3480
ggagcggcac cgcccagca gccccgcgc aggcgcgcgc cagccctcg cccgcaccc 3540
cgaccgcttc tcccacgaca gaactgcact tgtagccgga gacaactgta acctctctga 3600
tcggtttcac gaacacgaaa atggaaagtc ccggaaacgg agacacgaca gtgtggagaa 3660
cagtgacagt catgttgaaa agaaagcccc gaggagcgaa cagaaggatc ctctagaaga 3720
gcctaaagca aagaagcaca aaaaatcaaa gaagaaaaag aaatccaaag acaaacaccg 3780
agaccgcgac tccaggcatc agcaggactc agacctctca gcagcgtgct ctgacgtga 3840
cctccacaga cacaacaaaa aagaagaaga aaaagaagag acattcaaga aaatcagagg 3900
actttgttaa agattcagaa ctgcacttac ccagggtcac cagcttgagg actgtcgccc 3960
agttccggag agcccagggt ggctttcctc tctctggtgg cccgcctctg gaaggcgtcg 4020
gacctttccg tgagaaaacg aaacacttac ggtaggaaag cagggatgac aggtgtcgtc 4080
tctttgagta tggccagggt gattgaaaac tcagcctcaa aacaaaaaat tcactagtta 4140
tgattcaacg cgttcaacag aagccatccc cagcccagct taaattataa agatagacaa 4200
taactctgtt ccaatctgcg tgggtcttct ctgtacagat ttaccatgg 4260
agaacttttt ttttagtttt taccttttct taattacctt tattccgaat ggacgaacac 4320
tttctaccac tgctgaccat tgtaaaatac cgtgtatata aatccattg aaataatgcc 4380
ctggaataga acatctcaaa tgctgcttaa ttacagactc aggtcgatta cttgtatttc 4440
atgtaatgtt cctccaagtt agacatctgg tgcaagacca accgggagac catggaattg 4500
tcaaaagtac aaactgacag tgtgtatatt taatttaaag acttatttaa aaactcaca 4560
gctctcacct agactttgga gagcagctgt ttttctgtaa tgtctgatac tagaaactaa 4620
tttgcattat ttagttgtat tcaagatttg aagatgtat 4659

```

<210> 23

<211> 3711

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7484378CB1

<400> 24

```

atggagccca ctgtggctga cgtacacctc gtgccagga caaccaagga agtccccgct 60
ctggatgccg cgtgctgtcg agcgccagc attggcggtg tggccaccag ccttgtcgtc 120
ctcaccctgg gactcctttt gggaggaatg aacaactcca gacacgctgc cttaaagact 180
gcaacactcc ctgggaaggt ctacagcgtc actcctgaag caagcaagac cacgaaccca 240
ccagaaggaa gaaattccga acacatccga acatcagcaa gaacaaactc cggacacacc 300
atctttaaga aatgtaacac tcagcccttc ctctctacac agggcttcca cgtggaccac 360
acggccgagc tgcggggaat ccggtggacc agcagtttgc ggcgggagac ctcgactat 420
caccgcacgc tgacgcccac cctggaggca ctgctgcact ttctgctgcg accctccag 480
acgctgagcc tgggcctgga ggaggagcta ttgcagcgag ggatccgggc aaggctgcgg 540
gagcagggca tctccctggc tgcctatggc acaatttgtt acaatcaggcc gctgtccagg gaactccttt 660
cataagggac ctttggcaga aagagacttc aaatcaggcc gctgtccagg gaactccttt 660
tcctgcggga acagccagtg tgtgaccaag gtgaaccagg agtgtgacga ccaggaggac 720
tgctccgatg ggtccgacga ggcgcactgc gagtgtggct tgcagcctgc ctggaggatg 780
gccggcagga tcgtgggcgg catggaagca tccccggggg agtttccgtg gcaagccagc 840
cttcgagaga acaaggagca cttctgtggt gccccatca tcaacgccag gtggtggtg 900
tctgctgtc actgcttcaa tgagttccaa gacccgacga agtgggtggc ctacgtgggt 960
gcgacctacc tcagcggctc ggaggccagc accgtgcggg cccagggtgt ccagatcgtc 1020
aagcaccccc tgtacaacgc ggacacggcc gactttgacg tggctgtgct ggagctgacc 1080
agccctctgc ctttcggccg gcacatccag cccgtgtgcc tcccggtgc cacacacatc 1140
ttcccaccca gcaagaagtg cctgatctca ggctggggct acctcaagga ggacttccgt 1200
aagcatcttc ctcggcctgc aatggtcaag ccagaggtgc tgcagaaagc cactgtggag 1260
ctgctggacc aggcactgtg tgccagcttg tacggccatt cactcactga caggatgggt 1320
tgcgctggct acctggacgg gaaggtggac tctgccagg gtgactcagg aggaccctgt 1380
gtctgcgagg agccctctgg ccggttcttt ctggtggca tcgtgagctg gggaaatcggg 1440

```

```

tgtgcggaag cccggcgctcc aggggtctat gcccgagtca ccaggctacg tgactggatc 1500
ctggaggcca ccaccaaaagc cagcatgcct ctggccccc ccatggctcc tgcccctgcc 1560
gccccagca cagcctggcc caccagtcct gagagccctg tggtcagcac cccacacaaa 1620
tcgatgcagg cctcagtagc cgtgcctctt gactgggtca ccgttcctaa gctacaagaa 1680
tgtggggcca ggccctgcaat ggagaagccc acccggtcgt tggggcggtt cggagctgcc 1740
tccggggagg tgccctggca ggtcagcctg aaggaagggt cccggcactt ctgcggagca 1800
actgtggtgg gggaccgctg gctgctgtct gccgccact gcttcaacca cacgaagggt 1860
gagcaggttc gggccacact gggcactgct tccctcctgg gcctgggcgg gagcccgggt 1920
aagatcgggc tgcggcggtt agtgctgcac cccctctaca accctggcat cctggacttc 1980
gacctggctt tcctggagct ggccagcccc ctggccttca acaaatacat ccagcctgtc 2040
tgccctgccc tggccatcca gaagtccct gtgggcccga agtgcatgat ctccggatgg 2100
ggaaatacgc aggaagaaa tgccaccaag cccgagctcc tgcagaaggc gtccgtgggc 2160
atcatagacc agaaaacctg tagtgtgtct tacaacttct ccctcacaga ccgcatgatc 2220
tgcgcaggct tcctggaagg caaagtcgac tcctgccagg gtgactctgg gggccccctg 2280
gcctgcgagg agggccctgg cgtgttttat ctggcaggga tcgtgagctg gggatttggc 2340
tgcgctcagg ttaagaagcc gggcgtgtac acgcgcatca ccaggctaaa gggctggatc 2400
ctggagatca tgtcctccca gccccttccc atgtctcccc cctcgaccac aaggatgctg 2460
gccaccacca gcccaggac gacagctggc ctcacagtcc cggggggcac acccagcaga 2520
cccaccctg gggctgccag cagggtgacg ggccaacctg ccaactcaac ctatctgcc 2580
gtgagacca ctgctagggg acagacgcca ttccagacg ccccgagggc caccacacac 2640
accagctac cagactgtgg cctggcgccg gccgcgtca ccaggattgt gggcggcagc 2700
gcagcgggccc gtggggagtg gccgtggcag gtgagcctgt ggctgcggcg ccgggaacac 2760
cgttgcgggg ccgtgctggt ggcagagagg ttgctgctgt cggcgggcga ctgcttcgac 2820
gtctacgggg accccaagca gtggggcgcc tcctaggca cgccgttccct gagcggcgcg 2880
gaggggcagc tggagcgcgt ggcgcgcac tacaagcacc cgttctacaa tctctacacg 2940
ctcgactacg acgtggcgct gctggagctg gcggggcccg tgcgtcgcag ccgcctggtg 3000
cgtcccatct gccctgccga gccgcgcgg cgacccccgg acggcacgcg ctgcgtcatc 3060
accggctggg gctcgggtgc cgaaggaggc tccatggcgc ggcagctgca gaaggcgcc 3120
gtgcgcctcc tcagcgagca gacctgccgc cgcttctacc cagtgcagat cagcagccgc 3180
atgctgtgtg ccggcttccc gcagggtggc gtggacagct gctcgggtga cgctggggga 3240
cccctggcct gcaggagacc ctctggacgg tgggtgctaa ctgggggtcac tagctggggc 3300
tatggctgtg gccggcccca ctcccagggt gtctataccc ggggtggcagc tgtgagaggc 3360
tggataggac agcacatcca ggagtgacca ccacgtgact gccaggccg agactctacg 3420
tgaaagcaac aggagcagca ggccacccaa caccacgcg gccaccgtac cctacccaag 3480
gacgggtgtg ggggggctgt ggggtcatgg gatgcattct tgggtaccac cctttagttc 3540
caataaacac agcccctcca ccctagctca ctggctcagc acctcagtgt cacagcgaag 3600
gaccacatgc atggtgtctc accaggacct ggggtggcac taaggggaaa gatggacttc 3660
tcccaacca ggggaggctg agaccctccg agctgggggt ccagggacac g 3711

```

<210> 24

<211> 2017

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473143CB1

<400> 25

```

tgcagtgcaa gagtgtggca gatacaaagg acagaaacag gcaggatttg gctggaaagc 60
tgggtggatat gaagctggga ggtcactaag ggccaggcca tgtggaaagc cttgtatact 120
ttaattttcc ttattgaaga gcaaggagga gccattgaat gtttggggca tttgggaggt 180
tggcatgacc tcaccacctt ctgcgtgcag tgtgaagaac agattggcaa ggaccagagc 240
aaatgtggct gaccagttag gagttaatac ggcagtttag gaatgagctt ggtgtagggt 300
ggggacagac ggagatagag atagatgggt agattagcca tggggttgtg aagaagagga 360
agcttctagg tgagccttac ttagataaag agatggaggc atgattccat tctactgagtt 420
ggggggtagg caacagaaga ggagagagtg ggtgggggga catcgagagc atcccaaagg 480
ggtgatgggc ctggcccaca gagggatggc tggcctggat catgacgttg tgagtaacca 540
atgcacaagt gggaagtccc ccaaatcgga gagaggagca gaggccttgg cacggagact 600
gaaaggaggc agagaaagag caggagcagg aaaggagtat ggtattgtgg gaggaagctc 660
agggcattgc tgctcaaagt gtggtccac acagggcatc atcacatcac cagggagcat 720
ggtgggaagg cagtccctcc agctccaccc cggtgtcgat ctgaatctcc atttgagaca 780
gattccccag gtgatgcgtg tgcacagcca gaactgcacg ttccaactgc acggtcccaa 840

```

tgggacagtt	gagagcccag	ggttcccata	tggctacccc	aattacgcca	actgcacgtg	900
gaccatcacc	gcggaagagc	agcacagaat	ccagcttgtg	ttccagtcct	ttgccctgga	960
agaggacttt	gatgtectgt	cgggtgttga	tgggccaccc	cagccagaga	atctgcgtac	1020
gaggtctaca	ggctttcagc	tgccagccac	cattgttagt	gcagccacca	ccctctctct	1080
gcgcctcatc	agcgactatg	cagtcagtgc	ccaaggett	cagccacct	atgaagttct	1140
ccccagccac	acatgtggga	accagggag	gctgcccaat	ggcatccagc	agggttcaac	1200
cttcaacctc	ggtgacaagg	tccgctacag	ctgcaacctt	ggcttcttcc	tggagggcca	1260
cgccgtgtc	acctgccacg	ctggctctga	gaacagcgcc	acgtgggact	ttcccctgcc	1320
ttcctgcaga	gctgatgatg	cctgtggtgg	gacctgcgg	ggccagagt	gcatcatctc	1380
cagccccac	ttcccctcgg	agtaccataa	caatgccgac	tgacatgga	ccatcctggc	1440
tgaagctggg	gacaccatcg	ccctgggtgt	tattgacttc	cagctggagg	atgggttacg	1500
ctttctggaa	gtcactggga	cagaaggctc	ctccctctgg	ttcaccggag	ccagcctccc	1560
agcccccggt	atcagcagca	agaaactggc	gcgactgcac	ttcacatcgg	atggcaacca	1620
ccggcagcgc	ggattcagtg	ccaataacca	agtcaagaag	caaattgagt	tgaagtctcg	1680
aggtgtgaag	ctgatgcca	gcaaagacaa	cagccagaag	acgtctgtgt	gtttccacct	1740
cactcctcgt	gctgtctat	ctttgtcctc	tctgttgccg	tgtgtctaaa	tcctattagc	1800
tcagaaggtc	catgttcgat	gccacctctt	ccaggcagcc	tcacatgcgg	gtgcatcctt	1860
catccctccc	cactgtggtc	ccacagtccg	cttccgtgg	ttatgtctc	actcaactgg	1920
aaactccttg	aggacagtgg	tcttatctga	ctaccttcg	catttccatg	gtatccaaat	1980
aaagccttgt	acacagtaaa	aaaaaaaaaa	aaaaaaa			2017

<210> 25

<211> 2646

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4382838CB1

<400> 26

tccttctgga	tgttgtgggc	agaaagagta	cggccatctt	acagctgcat	tgccaataat	60
aatgtgggaa	accctgcaaa	aaagtccacc	aacatcattg	tgagagcatt	aaaaaaagga	120
cgatttttga	tcacaccaga	tccttatcac	aaagatgaca	acatccagat	tggccgtgag	180
gtgaaaatat	cttgccaagt	agaagctgtt	ccttctgagg	aggtaacatt	tagttgggtt	240
aaaaatggtc	gtccattaag	aagtctctg	cggatgggtc	ttacacagac	tgatcctgat	300
gtctctccgg	gaacaacaaa	cttggacatc	attgatttaa	aattcacgga	ttttgggacg	360
tacacatgta	tagcatctct	gaaggaggga	ggaatatctg	atatcagtat	cgatgttaat	420
atatccagtg	gcacagttcc	acccaatctg	actgttccac	aggaaaaatc	accattgggtc	480
accagagaag	gagacacaa	agaactgcaa	tgtcaagtaa	ctggcaaacc	taaaccaatc	540
atccttttgg	ctagagcgga	taaagaagtt	gcaatgcctg	atggatcaat	gcaaattggag	600
agttatgatg	gaacactgag	gattgtgaat	gtatctaggg	aaatgtcagg	aatgtacaga	660
tgtcagacca	ccaatacaaa	tggatttaac	gtgaaaccaa	gggaagcctt	ggtgcagctg	720
atcgttccgt	atccccctgc	agtggaaaca	actgttctgg	aaatccggca	aggacaggat	780
cgaagtgtca	ctatgagttg	cagagtactg	agagcctatc	caatacgggt	gctgacctat	840
gagtggcgct	tgggcaataa	attattacgg	acgggtcaat	ttgactctca	ggaatacaca	900
gagtacgctg	tgaagagtct	ttccaatgaa	aactatgggg	tttataactg	tagcatcata	960
aatgaagctg	gagctggggag	atgcagcttt	cttgtttacag	gaaaggccta	tgctccagaa	1020
ttctattatg	atactacaaa	tccagtatgg	cagaacagac	accgtgttta	ttctttacagt	1080
ctacagtggg	cacagatgaa	tcctgatgca	gtggatcgga	ttgttgcata	ccggttgggc	1140
atcaggcagg	ctggacagca	gcgctgggtg	gagcaggaga	ttaaaataaa	tgggaatatt	1200
caaaagggag	aattaattac	atataacttg	acagagctaa	ttaaaccaga	agcttatgaa	1260
gtccgactga	ctcctctcac	caaatttgg	gaaggagatt	caacaattcg	ggtgatcaaa	1320
tatagtgtct	ctgtaaatcc	tcatttgaga	gaatttcac	gtggatttga	agatggtaat	1380
atttatttgt	tcactcaaga	tgatacagat	aattttgact	ggacaaagca	aagtacagca	1440
acaagaaaata	caaaatatac	tcctaataca	ggacctaatg	ctgaccgtag	tggctccaaa	1500
gaaggttttt	atatgtacat	tgagacatca	cgaccagat	tgggaaggcga	aaaggctcga	1560
cttccagacc	ctgttttcag	catagctccc	aaaaaccctt	atggaccac	aaacactgca	1620
tattgtttca	gcttctttta	tcacatgtat	ggacaacata	taggtgtctt	aaatgtttat	1680
ctacgtttga	aagggcaaac	aacaatagag	aatccactgt	ggtcttcaag	tgggaataaa	1740
ggacaaagat	ggaatgaggc	tcatgttaat	atatacccaa	ttacttcatt	tcagctcatt	1800
tttgaaggta	tccgaggtcc	tggaaatagaa	ggtgacattg	ctattgatga	tgtatcaatt	1860
gcagaaggag	aatgtgcaaa	acaagaccta	gcaactaaga	attccgttga	tgggtgctgt	1920


```

gggatttttg ttcatatatg gctttttccc attatcgtcc tcatctctat cttaagtcct 1980
cgaaggtgac cttatcctgg cagaggctat aaaagattca ccaggcactg gcatgaagaa 2040
agagtctttg taaatggaca ttgaacaaac aaactaccaa agattcctcc actgactact 2100
gactcaaaaa taaaataata aaaacaaatt tttttaagcg ctggggataa aaagacatca 2160
tggaagtata acttattcca gactaaacat aaaagataat cttgacctga gtagagaaga 2220
gaccttcagg tgcttttgtg gctaaaaaga ttacagcgtc atctgggtga actctggaaa 2280
aaaaaaaaaa aaaaatgaaaa aaagaaaaaa aaaagagcta tagaaatcct tgtcaaagca 2340
caaagtcatg gctgggtttg tttcaaatga atagtgttct tgttaccatg gaaacctaat 2400
ggcctgccaa caaaaacctc actgtaaaaca gggtagctga agagctggca tttattttcc 2460
ttacgagaag gttttcgtag agaattaaat aaatgtaggc ccttttacct ttggctgtta 2520
cccttccttg aaaaataaac cgacttcgat ttttttaaag cttcctgttt tttaccacc 2580
tttttcccc tccccccctt attattatta ttattaatac cctggggtaa ggttgagtaa 2640
cataac

```

<210> 26

<211> 2088

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6717888CB1

<400> 27

```

atgggacctg cctgggtcca ggaccccttg acagggtgctc tctggctgcc tgtcctctgg 60
gcactcttgt cccaggctca ttgttttcat gaccaccag gatggcgctt cacttctca 120
gaaatttgtg tccccaggaa agtgcccac aggaggggtg gagttgagat gccagaccag 180
ctctcttaca gcatgcattt ccggggccaa agacacgtga ttcacatgaa gctcaagaag 240
aacatgatgc ccagacattt acctgttttt actaataatg accaaggggc catgcaggag 300
aactaccctt ttgtcccacg agactgttac tacgattgct acctggaagg ggttcctggg 360
tctgtggcca cattggacac ctgccgtgga ggtctgcgtg gcatgctgca ggtggatgac 420
ctgacttatg aaatcaaacc cctggaggct ttttccaaat ttgagtatgt agtatctctg 480
cttgtgtcag aagaaagacc aggagaggtc agtagatgta agactgaagg ggaagagata 540
gatcaagaat ctgaaaagg aaactggct gaaactccca gagaaggcca cgtttatttg 600
tggaggcatc atagaaaaaa cttgaaactt cactacacag ttactaatgg attattcatg 660
cagaacccta atatgtcaca cataatagag aatgtagtga ttattaacag catcatacat 720
accatttttc aaccagttta tttaaatgtc tatgtacgtg tttgtgcat atggaatgat 780
atggatatag taatgtataa catgcctgcc gacctgggtg taggagagtt tgggtcgtgg 840
aaatattatg aatggttttc acaaattcca catgatacct cagttgtttt tacatcaaat 900
cgacttgga acactcctcg ttgtggagac aagatcaaaa atcagaggga agaattgtgac 960
tgtggtctcc ttaaagattg tgccagtgat agatgttgtg agacctcttg taccctttct 1020
cttggcagtg tttgcaatac aggactttgc tgccataagt gtaaatatgc tgcccctgga 1080
gtggtttgca gagacttggg tggtatatgt gatctaccgg aatactgtga tgggaaaaag 1140
gaagagtgtc caaatgacat ctacatccag gatggaaccc catgttcagc agtatctgtt 1200
tgtataagag gaaactgcag tgaccgtgat atgcagtgtc aagccctttt tggctacca 1260
gtgaaagacg gttccccagc gtgctatcga aaattgaata ggattggtta ccgatttgga 1320
aactgtgggg ttattctacg gcgaggggga agtagacctt ttccatgtga agaagatgat 1380
gttttttgtg gaatgttgca ctgtagccgt gtcagccaca ttcccgggtg aggtgagcac 1440
actacatttt gtaatatatt agtacacgac ataaaagaag aaaaatgctt tggctatgaa 1500
gcacaccagg ggacagactt gccagaaatg gggctggtag tggatggtgc aacctgtggc 1560
ccaggagact actgtcttaa acgcaattgt actttttatc aagacctgca ttttgagtgt 1620
gatcttaaaa catgcaatta caaaggagta tgtaacaaca aaaacattg tcattgtctg 1680
catgagtggc aaccaccaac atgtgaactg agaggaaaag gaggtagtat agatagtggc 1740
cctctacctg acaacaata tcgtattgca ggcagcatac ttgtaaatac aaaccgagca 1800
ctagttttta tatgtattcg ttacatcctt tttgtgggtt cgcttctctt tgggtggctt 1860
tcacaagcaa tacaatgtta gggaagagaa aggaaagag cccacacatg gagtaaatga 1920
cattgacact tactgggaga tataatcaat agtcactctg acaattacat catcttttag 1980
caattctgat gtcattctga aataaaatcc cttggcaatt taaaaaggct tgtgtgttta 2040
aatttactta acatttcatg tctggtcaca ttctcaatac ttctatag

```

<210> 27

<211> 1890

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472044CB1

<400> 28

atgctgctgg	ctgtgctgct	gctgctaccc	ctcccaagct	catggtttgc	ccacggggcac	60
ccactgtaca	cacgcctgcc	ccccagcgcc	ctgcaagtct	tcactctcct	cttggggggca	120
gagactgtgt	tggggcgcaa	cctagactac	gtttgtgaag	ggccgtgcgg	cgagaggcgt	180
ccgagcactg	ccaatgtgac	gcggggccac	ggccgcacgc	tggggggcag	cgcgggcgccg	240
cccggggcct	ggccctggct	ggtgaggctg	cagctcggcg	ggcagcctct	gtcgggcggc	300
gtcctggtag	cggcctctcg	ggtgctcacg	gcagcgcact	gctttgtagg	ctgccgctcg	360
acccgcagcg	ccccgaatga	gcttctgtgg	actgtgacgc	tggcagaggg	gtcccggggg	420
gagcaagcgg	aggaggtgcc	agtgaaccgc	atcctgcccc	acccaagtt	tgaccgcg	480
accttcacac	acgacctggc	cctggtgcag	ctgtggacgc	cggtgagccc	gggggggatcg	540
gcgcgccccg	tgtgcctgcc	ccaggagccc	caggagcccc	ctgccggaac	cgccctgcgc	600
atcgcgggct	ggggcgccct	cttcgaagac	gggcctgagg	ctgaagcagt	gagagaggcc	660
cgtgttcccc	tgctcagcac	cgacacctgc	cgaagagccc	tggggcccgg	gctgcgcccc	720
agcaccatgc	tctgcgcggg	gtacctggcg	gggggcgttg	actcgtgcc	gggtgactcg	780
ggaggcccc	tgacctgttc	tgagcctggc	ccccgcctta	gagaggctct	gttcggagtc	840
acctcctggg	gggacggctg	cggggagcca	gggaagcccc	gggtctacac	ccgcgtggca	900
gtgttcaagg	actggctcca	ggagcagatg	agcgctcctc	cctccagccg	cgagcccagc	960
tgcagggagc	ttctggcctg	ggaccccccc	caggagctgc	aggcagacgc	cgccccggctc	1020
tgcgccttct	atgcccgcct	gtgcccgggg	tcccagggcg	cctgtgcgcg	cctggcgcac	1080
cagcagtgcc	tgacgcgcg	gcggcgatgc	gagctgcgct	cgctggcgca	cacgctgctg	1140
ggcctgctgc	ggaacgcgca	ggagctgctc	gggcctcgtc	cgggactgcg	gcgcctggcc	1200
cccgccttgg	ctctccccgc	tccagcgctc	agggagctct	ctctgcaccc	cgccccgggag	1260
ctgcggcttc	actcaggctg	ccctgggctg	gagcccccgc	gacagaagtt	ggctgcccctg	1320
cagggggccc	atgcctggat	cctgcaggtc	ccctcggagc	acctggccat	gaactttcat	1380
gaggtcctgg	cagatctggg	ctccaagaca	ctgacccggc	ttttcagagc	ctgggtgcgg	1440
gcaggcttgg	ggggccggca	tgtggccttc	agcgccctgg	tgggcctgga	gccggccaca	1500
ctggctcgca	gcctcccccg	gctgctgggt	caggccctgc	aggccttccg	cgtggctgcc	1560
ctggcagaag	gggagcccca	gggaccctgg	atggatgtag	ggcagggggc	cgggctggag	1620
aggaaggggc	accacccact	caaccctcag	gtaccccccg	ccaggcaacc	ctgagccatg	1680
tctggggccc	cagcccctgg	ggaggacctc	ctgctcccag	gggctgagag	gggttcggga	1740
gcataatgac	aaactgtcgc	tgccccagtg	gctgggtgtg	tgtgggtggg	atgggggtggg	1800
ggtcctgggc	cccccgctgc	ttcccaggtt	tacaatcaga	gaatcacagc	tgctttaata	1860
aatgttattt	ataataaaaa	aaaaaaaaaa				1890

<210> 28

<211> 2984

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477384CB1

<400> 29

agtgaagacg	actgtcttat	ctgactgtag	gcacagagag	tgcgccgcga	gagggcggtct	60
cctcacgcgtc	aggcgccggc	aggctgcggt	ctctgctggc	cgacgcccga	aggcgccgaa	120
tggggggggc	ctgccgagct	cccttacagc	cccaatgtgc	gcgcgcgcgg	gaggcttggg	180
cacgcaggca	ccgcgcggcg	ggggcggggc	gaaggcgggc	gggcggggca	ccagctgcgc	240
gcgcggggcg	ggggcggggc	cgggggcggg	gcgcgctgct	tgggtcccgc	cgccctggg	300
ctcctcccc	tcccgcgccc	aggccagcgg	cgggcccagc	tctcccccg	actcgggtctc	360
tctccccctc	cctccgcgcc	gcagttcctc	cctcccgccg	ccgcctcttc	ctcgggtgagg	420
cgctcttcca	gcgggcaggc	agcatggcgg	ccgtggagac	gcgggtgtgc	gagacagacg	480
gctgcagcag	tgaggccaag	ctccagtgtc	ccacttgcat	caagctgggc	atccagggct	540
cgtacttctg	ctcgcaggaa	tgttttaaa	gaagttgggc	tactcacaag	ttactacata	600
agaaagcaaa	agatgaaaag	gcgaagcgag	aagtgtcttc	ctggactgtg	gaagggtgata	660
ttaatactga	cccatgggca	ggttatcgat	atactggtaa	actcagacca	cattatccac	720
tgatgccaac	aaggccagtg	ccaagttata	ttcaaagacc	agattatgct	gatcatccct	780

```

taggaatgtc tgaatctgaa caggtcttta aaggtacttc tcagattaaa ttactctcat 840
ctgaagatat agaagggatg cgacttgtat gtaggcttgc tagagaagtt ttggatgttg 900
ctgccggcat gattaaacca ggtgttaacta ctgaagaaat agatcacgct gtacacttag 960
catgtattgc aagaaattgc tacccttctc ccttgaatta ttataatttc ccaaagtctt 1020
gttgtacctc agtgaatgaa gtcatttgcc atggaatacc agacagaagg cccttacaag 1080
aaggtgacat tgttaatgtg gatatcactc tttatcgcaa tggttatcat ggggacctga 1140
atgagacatt ttttgttgga gaagtggatg atggagcacg gaaacttgtt cagaccacat 1200
atgagtgcct gatgcaagcc attgatgcag tgaagcctgg tgttcggtac agagaattgg 1260
gaaacattat ccagaagcat gcccaagcaa atgggttttc agttgttcga agctattgtg 1320
ggcatggaat ccacaagctt tttcatacag ctcccaatgt accccactat gctaaaaata 1380
aagcagttgg agtgatgaag tcggggccatg tatttacaat tgagccaatg atttgtgaag 1440
gcggatggca ggatgaaacc tggccagatg gttggactgc ggtgacaaga gacggaaagc 1500
ggtctgctca gtttgagcac accctcctgg tcacagacac tggctgtgaa atcctaacc 1560
ggcgacttga cagtgcacgg cctcacttca tgtctcaatt ttaatttctc ccaagatggc 1620
acattctcag accttcttac tgtgctatgc attttattga gagtacagaa aggaagagga 1680
accttttttt aatcacttgt tttgttttga ctatagataa gaaaggacta cagcatttga 1740
tgtgtgtcct caagaacttg tcttgggtct gaaaaagctg agaagaataa aggaaacatt 1800
gctcaactct tcagccccct cccctgcac acctgttttc tcatttgccc tttgagcact 1860
tttacttaaa cttgcttgta gttgctttta tcaactgccg aaaacagcca tcaagagcca 1920
tctgctttcc aggtgaacat tggaaatgag aatctttgaa acttagcaat atgtgttgca 1980
ccagattttt taaattatat atatggaaat atatatgtat acattttaag ttctgtatac 2040
ataattacca aacactatgt gacctggagt ttgtgttgtt tctgctctga cagggttata 2100
tgttctttaca aatggatcca tagtttgtag tgatttaatt cctggttggg atttggcctc 2160
ccctctcccc catgctaatt atttaccctt gtaatttgtc ataggggaagc actcacccaa 2220
tgagactttc tccaatgtgg actctgtgtg tcagtgaatg aatgtagtaa aattcacttt 2280
ggaaggttat caggctttta aaaatctagt ttatggcaaa aatagccatt ttccaagtgg 2340
tggtgactg ttgcagggaa tgagaatttc ataatacact gctatttcag acctctgttt 2400
ggtcagaaat ggaaaagaaa aagccccctt tcttcccttt tctgttttac ttcaagggca 2460
taccttgagg gtgctcagag aagcgtgaag tttgcactat ggtggaggat ggggaaagag 2520
ttctaaagtg tctccagctg tgaacccagg aggtcaagtg ggctattaaa atctaactgt 2580
gagtaaagtg gatagtgat agaaaggaat tttgtgtact gtaaccttgc agtagagatg 2640
cagctgtcct tcgtgtgtgg aaacacacct ctcccttaca tagttgggaa cctcattaga 2700
aatgacctca gctgccccat atctacgttc ctttcagcag ttgtccaagt aggagtgtat 2760
ccagtgaaga catatcaaat cacaaagtca ttgtcattag agtgacttg attactgggc 2820
atccttgtaa tataatttca taccactgac acattatact tgtaagagaa catctttccc 2880
agagtgcctc agaccttatt gctttaaaat ataataatgt tttcattact tttattattt 2940
gaatgattta gtaaagtga ctgaatctgg ttagactttt ggga 2984

```

<210> 29

<211> 2255

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7077175CB1

<400> 31

```

ccacagtgtg gatgcccctt gaggatgtca cactcatgag acgccagaca caaaacgcca 60
cacagtgtgt aatcccattt ccatgaaatg tccaggtcag gccagtgcac agacacaggc 120
agcgggtgtg tgggcagcgg ggctggagaa ggggacgggg agtgaccgct gagggggaca 180
ggcttctttt agtggggatg aacgttctaa aattggacac attggtgttg gcacagctgt 240
ggagatatga aaacgcgaaa cctacgggtg agctgggtga acctatgag gcgggaatta 300
actgctccgg ctctggcgct gaggaaaagg aggaccggag gatggcgatc atctgggccg 360
tgccctccac atctgtgtcc tgggaacaga cttctagaaa aacccaaatc agggaaaaagc 420
ggccagctcc acgctgcaa cagctgggca ccaggcagag agtgttacca gtggtcaagc 480
cagaggtgct gcagaaagcc actgtggagc tgctggacca ggcactgtgt gccagcttgt 540
acggccattc actcactgac aggatggtgt ggcgtggcta cctggacggg aaggtggact 600
cctgccaggg tgactcagga ggaccctgg tctgcagga gccctctggc cgggtcttct 660
tggctggcat cgtgagctgg ggaatcgggt gtgcggagc ccggcgcca ggggtctatg 720
cccaggtcac caggctacgt gactggatcc tggaggccac caccaaagcc agcatgcctc 780
tggcccccac catggctcct gccctgccc cccccagcac agcctggccc accagtcctg 840
agagccctgt ggtcagcacc cccaccaaat cgatgcaggc cctcagtacc gtgcctcttg 900

```

actgggtcac	cgttcctaag	ctacaagaat	gtggggccag	gcctgcaatg	gagaagccca	960
ccgggggtcgt	gggcggttc	ggagctgcct	ccggggaggt	gccctggcag	gtcagcctga	1020
aggaaggggtc	ccggcacttc	tgcggagcaa	ctgtggcggg	ggaccgctgg	ctgctgtctg	1080
ccgcccactg	cttcaaccac	acgaagggtg	agcagggttcg	ggcccacctg	ggcactgcgt	1140
ccctcctggg	cctgggcggg	agcccgggtga	agatcggggt	gcggcgggta	gtgctgcacc	1200
ccctctacaa	ccctggcatc	ctggacttctg	acctggctgt	cctggagctg	gccagccccc	1260
tggccttcaa	caaatacatc	cagcctgtct	gcctgcccct	ggccatccag	aagttccctg	1320
tggggccggaa	gtgcatgac	tccggatggg	gaaatacgca	ggaaggaaat	gccaccaagc	1380
ccgagctcct	gcagaaggcg	tccgtgggca	tcatagacca	gaaaacctgt	agtgtgctct	1440
acaacttctc	cctcacagac	cgcatgatct	gcgcaggctt	cctggaaggc	aaagtgcact	1500
cctgccaggg	tgactctggg	ggccccctgg	cctgcgagga	ggccccctggc	gtgttttatc	1560
tggcagggat	cgtgagctgg	ggtattggct	gcgctcaggt	taagaagccg	ggcgtgtaca	1620
cgcgcatcac	caggctaaag	ggctggatcc	tggagatcat	gtcctcccag	ccccttccca	1680
tgtctcccc	ctcgaccaca	aggatgctgg	ccaccaccag	ccccaggacg	acagctggcc	1740
tcacactccc	gggggcccac	cccagcagac	ccacctctgg	ggctgccagc	agggtgacgg	1800
gccaacctgc	caactcaacc	ttatctgccg	tgagcaccac	tgctagggga	cagacgccat	1860
ttccagacgc	cccggaggcc	accacacaca	cccagctacc	aggtagccgg	agagacggag	1920
ggatccctgg	gagtggaggg	tcccatgtta	atcagcctgg	gctgcctaac	aagacataac	1980
gtcgtccact	ttggggaggc	gaggcgggcg	gatcaagagg	tcaggagatc	gagaccatcc	2040
tggcgaacac	ggtgaaacct	tgtctctact	aaaaaatac	aaaaaattag	ccaggcgtgg	2100
tgggtgggcg	ctgtagtccc	aactacgcgg	gagggttaagg	caggagaatg	gcatgaagcc	2160
gggaggcgga	gcttgcatg	agctgcatgc	cactgcactc	cagcctggca	acaagcgaaa	2220
ctcctgtctca	aaaaagaaaa	agacataacg	gcctc			2255

<210> 30

<211> 1250

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7480124CB1

<400> 32

ccgtcatggg	cccactcggg	ccctctgccc	tgggccttct	gctgctgctc	ctggtggtgg	60
cccctccccg	ggtcgcagca	ttgggtccaca	gacagccaga	gaaccaggga	atctccctaa	120
ctggcagcgt	ggcctgtggt	cggcccagca	tggaggggaa	aatcctgggc	ggcgtccctg	180
cgccccagag	gaagtggccg	tggcaggtca	gcgtgcacta	cgcaggcctc	cagctctcg	240
gcggctccat	cctcaatgag	tactgggtgc	tgctcagctgc	gcactgcttt	cacagggaca	300
agaatatcaa	aatctatgac	atgtacgtag	gcctcgtaaa	cctcaggggtg	gccggcaacc	360
acacccagtg	gtatgggggtg	aacagggtga	tcctgcaccc	cacatatggg	atgtaccacc	420
ccatcggagg	tgacgtggcc	ctggtgcagc	tgaagacccg	cattgtgttt	tctgagtccg	480
tgctcccggg	ttgccttgca	actccagaag	tgaaccttac	cagtgccaat	tgctgggcta	540
cgggatgggg	actagtctca	aaacaagggtg	agacctcaga	cgagctgcag	gaggtgcagc	600
tcccgtgat	cctggagccc	tgggtgccacc	tgctctacgg	acacatgtcc	tacatcatgc	660
ccgacatgct	gtgtgctggg	gacatccctga	atgctaagac	cgtgtgtgag	ggcgactccg	720
ggggcccact	tgtctgtgaa	ttcaaccgca	gctgggttga	gattggaatt	gtgagctggg	780
gccgaggctg	ctccaacct	ctgtaccctg	gagtgatgc	cagtgtttcc	tattttctcaa	840
aatggatatg	tgataacata	gaaatcacgc	ccactcctgc	tcagccagcc	cctgctctct	900
ctccagctct	ggggcccact	ctcagcgtcc	taatggccat	gctggctggc	tggtcagtgc	960
tgtgaggtca	ggataccac	tctaggattc	tcatggctgc	acaccctgcc	ccagcccagc	1020
tgctccaga	cccctaagca	tctcctgtcc	tggcctctct	gaagcagaca	agggccacct	1080
atccccgggg	tggatgctga	gtccaggagg	tgatgagcaa	gtgtacaaaa	gaaaaagggg	1140
aagggggaga	ggggctggtc	agggagaacc	cagcttgggc	agagtgcacc	tgagatttga	1200
taagatcatt	aaatatttac	aaagcaaaaa	aaaaaaaaaa	aaaaaaattg		1250